A project to

Develop and
Evaluate
Methods to
Promote
Ambulatory Care
Quality

Final Report, Volume 1

RA
399
.A3
D46
1994
v.1

DEMPAQ:
A project to Develop and Evaluate Methods to Promote Ambulatory Care Quality

Final Report, Vol. 1

This report describes the development and evaluation of a set of performance measures designed to assess the quality of care delivered to Medicare beneficiaries. Because this project was a test of a methodology, the results presented in this report should be regarded as preliminary rather than definitive.

R. Heather Palmer, M.B., B.Ch., S.M.

Linda E. Clark, R.N.

Ann G. Lawthers, Sc.D.

Jean E. Edwards, B.A.

Jinnet Fowles, Ph.D.

Deborah Garnick, Sc.D.

Jonathan Weiner, Dr.PH

1994

Table of Contents, Volume 1 and 2

Chapter 1
Conclusions and Recommendations

Chapter 2
Overview and Philosophy of the DEMPAQ Project ✓

Chapter 3
Building a Collaborative Project ✓

Chapter 4
Focus on Quality: Profiling Physicians' Practice Patterns ✓

Chapter 5
Developing a Quality Improvement Database Using Health Insurance Data: ✓
A User's Guide with Application to Medicare's National Claims History File

Chapter 6*
The DEMPAQ Claims Review System: Results for Maryland

Chapter 7*
The DEMPAQ Claims Review System: Results for Alabama

Chapter 8*
The DEMPAQ Claims Review System: Results for Iowa

Chapter 9
Agreement Between Physicians' Office Records and Medicare Claims Data

Chapter 10
Evaluating the DCRS: Assessment by Physicians

Chapter 11
The Record Review Instrument

Chapter 12
Comparison of Participants and Nonparticipants in the DEMPAQ Record Review

Chapter 13
Evaluating DRRS: Inter-rater Reliability for Data Abstraction

Chapter 14
Evaluating the DRRS: Assessment by Physicians

Chapter 15
Evaluating DRRS and DCRS: Feedback of Results to Physician Groups

** Contained in Volume II*

Chapter 16*
The DEMPAQ Record Review System: Results for Maryland

Chapter 17*
The DEMPAQ Record Review System: Results for Alabama

Chapter 18*
The DEMPAQ Record Review System: Results for Iowa

Chapter 19
Variation in Office-Based Quality:
A Three-State Comparison of Care Provided to
Medicare Patients with Diabetes and Congestive Heart Failure

Chapter 20
Patterns of Variation, Record Review

Chapter 21
Costs of Implementing the DEMPAQ Record Review System (DRRS)

References

Table of Contents, Volume 3

Chapter 1
Summary and Recommendations

Chapter 2
Background and Overview of DEMPAQ Feedback and Education

Chapter 3
The CME Workshops

Chapter 4
CME Workbook

Chapter 5
Claims Reports

Chapter 6
Records Reports

References

** Contained in Volume II*

Acknowledgments

The DEMPAQ project involved the collaboration of many individuals in diverse geographic locations from different institutional affiliations.

Peer Review Organizations

Delmarva Foundation for Medical Care, Inc. (DFMC)
Special acknowledgment is being made to Linda E. Clark, R.N., who in her role as Chief Executive Officer and Executive Vice President of the Delmarva Foundation for Medical Care, Inc., recognized the need for the Foundation to engage in research in the area of ambulatory care quality. She reached out to the health care research community through the American Medical Review and Research Center, and, together with Dr. Heather Palmer, was successful in obtaining funding from the Health Care Financing Administration for this project that we now call DEMPAQ.

The Delmarva Foundation for Medical Care has been the prime contractor for the project. At DFMC, Jean Edwards has acted as the Project Director and coordinated many of the administrative aspects of the project. She represented the PRO community during researcher meetings and conference calls and ensured that project deliverables and deadlines were met.

Sandra Ray Moentmann, M.S., R.N., Project Manager, coordinated the record review for DFMC and the two other PROs. She developed materials with the research team to guide the review and trained and monitored the review teams in Alabama and Iowa. She also coordinated the physician review of records and claims in Maryland and directed the claims reliability study. Ms. Moentmann completed the final proofing of this report. Finally, Ms. Moentmann has had an active role in the Feedback and Education Modification to this project.

Lori Collier, R.N., served as review supervisor. Under her able guidance, the reviewers completed the record review study in Maryland and initiated the physician review of records. The nurse reviewers at DFMC included Cathy Asche, Marge Boone, Judy Brower, Nancy Eber, Geri Gill, Janet Herlihy, Teri Richardson, and Leona Schmidt.

Jill Jenkins, Publications Manager, designed and produced this report, the *DEMPAQ Quarterly*, and the overheads and slides for the DEMPAQ project. She also designed the CME workbook and the slides used for the CME activity. Mary Ann Cowley, Publications Assistant, assisted with the production of this report.

Secretarial support staff for the project included: Carrie Cassidy, Executive Secretary; Bevlee Burks, Assistant Executive Secretary; and Annette Newcomb, Assistant Executive Secretary. This staff typed monthly reports and deliverables, as well as completed the transfer of documents from the researchers and organized the attachments for this report.

Cherie Adams, Cleora Wheedleton, and Danielle Helmer were responsible for copying the two reports to physicians (claims and records) and the CME workbook. As each report book was tailored to the individual, this difficult task involved producing almost 300 unique books. Ms. Adams, Ms. Wheedleton, and Ms. Helmer are also responsible for the copying of this final report.

Seven physicians recruited by the primary specialty societies and trained by DFMC participated in the peer review activities of the project. These individuals were Joseph Zebley, III, M.D.; Sheldon Goldgeier, M.D.; Christine Marino, M.D.; Alfred Ossman, M.D.; Howard Wilson, M.D.; Abdul Nayeem, M.D., and Christian E. Jensen, M.D., M.P.H.

Alabama Quality Assurance Foundation, Inc. (AQAF)
Nadine Schiesz, R.N., was Alabama's primary liaison to the DEMPAQ project. She coordinated all aspects of AQAF's participation in the project: the record review; the liaison briefing sessions; the CME workshop; and periodic project meetings.

Laurie Hall, M.D., physician reviewer at AQAF, conducted all of the peer review for AQAF and contributed substantially to the CME workshop in Alabama.

Carlene Gibbons, R.N., and Barbara Baites, R.N., were the nurse reviewers for Alabama.

Iowa Foundation for Medical Care, Inc. (IFMC)
Nancy Paine, R.N., was the primary liaison for the Iowa Foundation for Medical Care to the DEMPAQ project. She coordinated all aspects of IFMC's participation in the project: the record review; the briefing sessions; and the CME workshop. Ms. Paine was assisted by Sherrie Segebart.

Jenny Van Gundy, R.N., Dottie Klein, R.N., and Nina Larson were nurse reviewers for Iowa.

Thirteen physicians, recruited by the primary specialty societies and trained by IFMC, participated in the peer review activities of the project. These individuals were Milton VanGundy, M.D.; Steven Craig, M.D.; Robert Good, D.O.; Kevin deRegnier, D.O.; John Carroll, M.D.; Jay Rosenberger, D.O.; Darwin Shassow, D.O.; Kevin Quinn, M.D.; Larry Severidt, M.D.; Mary Ann Nelson, M.D.; Andy Edwards, M.D.; Lane Dvorak, M.D.; and Dana Shaffer, D.O.

Researchers
R. Heather Palmer, M.B., B.Ch., S.M., of the Harvard School of Public Health (HSPH) has been the project's principal investigator. She provided the core concept for the overall project.

Ann G. Lawthers, Sc.D., also from HSPH, served as the project's Research Director. Dr. Lawthers contributed to, and assumed responsibility for, many aspects of the project, both record and claims. She was the critical technical link between the PROs and the researchers, and coordinated many of the CME activities for the Feedback and Education Modification. She assumed a major role in the completion of this final report and authored six chapters relating to the DEMPAQ Record Review System (DRRS).

Jinnet Fowles, Ph.D., of Park Nicollet Medical Foundation, coordinated the claims profiling effort and assumed primary responsibility for the design of claims-record comparison and the participant/nonparticipant analysis.

Deborah W. Garnick, Sc.D., of the Heller School of Public Policy, Brandeis University, contributed to the claims profiling and designed and executed the cost study.

Jonathan Weiner, Dr.PH of the School of Hygiene and Public Health, Johns Hopkins University, contributed to the claims profiling component of the project and directed the Johns Hopkins team responsible for claims profiling systems development and data processing.

Many individuals at the research institutions contributed to the DEMPAQ project. At Johns Hopkins, Stephen Parente, M.S.P.H., was the project manager of the claims profiling activity. Other members of the systems analysis/programming team at the Johns Hopkins Health Services Research and Development Center included Tom Richards, M.S.E., Paul Chandler, and Andrew Baker.

At the Harvard School of Public Health, Naomi Banks, M.B.A., designed the record review system and the peer review of records. Howard Frazier, M.D., provided consultation and clinical recommendations for development of the record review criteria. John Orav, Ph.D., provided statistical direction for both the record and the claims studies. The following consultants, Gerry Schumacher, D.Pharm., Patricia O'Shea, M.D. and Helen Burstin, M.D., drafted the initial criteria for the project. Mary Adams, Ph.D., Mark Monane, M.D., and Carol Ellenbecker, Ph.D., researched the references for the criteria and prepared detailed documentation of the clinical rationales. Barbara Lightfoot typed countless tables for feedback to physicians, both liaison and participant.

At Park Nicollet Medical Foundation, Doris Petrie performed the programming and statistical analyses for the claims reliability study and the participant/nonparticipant analysis.

Data Consultants

Several individuals gave of their time to review the claims profiling specifications (see Chapter 10). These individuals included: Neil Powe; Lisa Iezzoni; Hal Luft; Elliot Fisher; and Kathleen Lohr.

Liaison Physicians

The DEMPAQ project has relied heavily on the good will and input from practicing physicians. Many individuals at both the national and state level have devoted countless unpaid hours to the review and development of the DEMPAQ criteria.

National Consultants	
American Osteopathic Association	James Cole, D.O. Director of Professional Affairs Cranston General Hospital Cranston, Rhode Island
American Academy of Family Physicians	Harry L. Metcalf, M.D. Williamsville, New York
American College of Physicians	G.C. Woodson, M.D. Decatur, Georgia
American Medical Association	W. Lamar Weems, M.D. Jackson, Mississippi
American Society of Internal Medicine	William E. Golden, M.D. Director, General Internal Medicine Division University of Arkansas for Medical Sciences Little Rock, Arkansas
Maryland	
Medical-Chirurgical Faculty of Maryland	Robert Ruderman, M.D.
Maryland Academy of Family Physicians	Joseph Zebley, M.D.
Maryland Society of Internal Medicine	Sheldon Goldgeier, M.D. Park Espenschade, M.D.
Alabama	
Medical Association for the State of Alabama	Peter Morris, M.D.
Alabama Chapter of the American Society of Internal Medicine	Pink Folmar, M.D.

Alabama Chapter of the American Association of Family Physicians	Brian Perry, M.D.
Iowa	
Iowa Medical Society	Jeffrey Ver Huel, M.D.
Iowa Clinical Society of Internal Medicine	Steven Craig, M.D. Mary Stuart, M.D.
Iowa Academy of Family Physicians	Milton VanGundy, M.D.
Iowa Osteopathic Medical Association	Robert G. Good, D.O. Kevin deRegnier, D.O. Gary Janssen, D.O.

Committees

Several committees were convened at the beginning of the DEMPAQ project to facilitate communication among the diverse parties and groups involved.

Scientific Oversight Committee

This committee reviewed the initial project design. Members included: William E. Golden, M.D.; Robert L. Kane, M.D.; Kathleen N. Lohr, Ph.D.; Carole Magoffin, M.S.; Michael R. McGarvey, M.D.; David B. Nash, M.D., M.B.A.; and J. William Thomas, Ph.D.

Project Committee

The Project Committee was formed to provide overall direction to the project, to provide group problem identification and resolution, and to maintain various liaison relationships. The Committee met during the project's early phases. It consisted of Linda Clark (DFMC), Heather Palmer (HSPH), Jean Edwards (DFMC), Ann Lawthers (HSPH), Nancy Paine (IFMC) and Nadine Schiesz (AQAF).

PRO Advisory Committee

The PRO Advisory Committee consisted of individuals representing the original seven PROs interested in the DEMPAQ project: Heather Palmer (HSPH); Linda Clark (DFMC); Jean Edwards (DFMC); Dan Emmott (UPRO); Elizabeth Hamel (Health Care Review, Inc.—Rhode Island); Nadine Schiesz (AQAF); Marcia Petrillo (ConnPRO); Debra Nixon (Health Services Advisory Group, Inc.); and Frederick Dettmann (WPRO). As with the other committee, the PRO Advisory Committee met only during the project's early phases.

Record Review Committee

This committee was designed to review and advise on the methods and criteria for the record review. It met during the first year of the project and consisted of Heather Palmer (HSPH), Linda Clark (DFMC), Jean Edwards (DFMC), Christian Jensen (DFMC), James Maroc (IFMC), Robert Sherrill (AQAF), and Ann Lawthers (HSPH).

Health Care Financing Administration

Paul Elstein, Ph.D., initially served as the project officer and assisted the project in many ways. He helped to secure the required databases, established liaisons between HCFA and the research team, and provided continuous guidance throughout the three years.

Peggy Bowen, R.N., served as the project officer for the latter part of the project and provided support and guidance to the project and in particular to DFMC, the prime contractor.

Executive Summary

DEMPAQ stands for *D*eveloping and *E*valuating *M*ethods to *P*romote *A*mbulatory care *Q*uality. The project, which used Medicare claims and physician office records to evaluate the quality of care given to Medicare beneficiaries in primary care physicians' offices, was initiated in fall 1990 and completed in December 1993. In funding the project, the Health Care Financing Administration (HCFA) agreed to allow us to test and evaluate techniques that could be used to improve the quality of primary care for Medicare beneficiaries.

Our approach to improving quality is built upon the principles of continuous quality improvement. A successful quality improvement design has the following components:

Step 1—achieve consensus about parameters to measure quality;
Step 2—measure performance;
Step 3—identify targets for improvement based on the results of the measurement;
Step 4—intervene to change practice; and
Step 5—re-measure performance to assess the impact of the intervention.

During the DEMPAQ project, we worked closely with general medicine physician societies that acted as our liaisons with national and state physicians to identify relevant national practice guidelines and policies (step 1). To accomplish step 2 we built clinical performance measurement tools and evaluated their reliability and validity in three states: Maryland, Alabama and Iowa. A feedback and education modification to the original project tested the feasibility of undertaking steps 3 and 4 in the same three states. A single intervention, an integrated CME program, was chosen as the test intervention. Step 5 awaits future funding.

In this document, we present an overview of the performance measures we developed and discuss the fundamental principles that guided our work.

Overview

We built two complementary sets of tools: one set based on Medicare claims data and one set based on office records. Both the claims and records tools were subject to rigorous reliability and validity testing using a structured physician review instrument.

Claims-based profiles

The claims tools included:

- condition-specific profiles;
- office practice profile; and
- prototypical preventive services profile.

Condition-specifc profiles were developed for patients with diabetes, hypertension, congestive heart failure, ischemic heart disease, chronic obstructive pulmonary disease, and osteoarthritis. Each condition-specific profile included four categories of performance measures:

- services recommended for all patients with the condition;
- services which are either provided infrequently or only to a subset of patients with the condition;
- complications which may result from inadequate ambulatory care and require hospitalization; and
- use of specialists.

The office practice profile portrays the care for patients of a given primary care physician, including care given by all other providers, with adjustments for casemix using ambulatory care groups, and using resource value units to summarize all services used by these patients.

Records-based profiles

We designed the records-based profiles around "functions," that is, activities that physicians typically perform in the course of an office visit:

- making and monitoring a ***diagnosis;***
- responding to an abnormal ***test*** value;
- prescribing and monitoring a ***drug;***
- performing a ***procedure;***
- ***following-up*** serious health threats;
- rendering ***preventive*** or screening services; and
- ***documenting*** in the medical record.

Unique performance measures were constructed for individual tests, diagnoses, drugs, procedures and preventive services.

Basic Principles

We constructed the DEMPAQ performance measures on the following three basic principles:

- *Collaborative Efforts*—the process for developing the measures should be interactive and open, using a dialogue between user and developer;
- *Performance Measures*—the content of performance measures should be clinically useful and focus on the average quality of health care received by beneficiaries in a state; and
- *Feedback Results*—the measures should be designed for education and feedback to improve the quality of care for all physicians rather than for tagging physicians as outliers and initiating sanctions.

The following describes these principles in more detail.

Collaborative Efforts

Working relations

The DEMPAQ project represented a cooperative venture among the Peer Review Organization (PRO), research and physician communities. The Delmarva Foundation for Medical Care, Inc., the PRO for Maryland, collaborated with researchers at the Harvard School of Public Health, Brandeis University, Johns Hopkins University and Park Nicollet Medical Foundation in Minneapolis, as well as with the PROs in Alabama and Iowa, to design and test quality measurement tools.

The cornerstone of our measure development has been a "customer-supplier" dialogue between user (physicians) and developer (PROs and researchers). Physicians have been involved in all phases of the project, development, implementation, evaluation and feedback. By beginning an interactive relationship with physician groups during the initial development of measures, we reached agreement, from the start, upon performance measures that met the physicians' own sense of good practice and that were clinically relevant. As a result, well developed working relations between physicians and PROs and between physicians and researchers evolved over the course of the project. Each of the states involved in the project developed unique styles of interaction, appropriate for their own environment.

By the end of the project, the PRO-researcher relationship became similarly well developed. PROs had the opportunity to tap researcher expertise while researchers learned practical implementation lessons.

Spheres of influence

The points of contact between the project and the physician community varied in each state. However, the common theme was that physician-to-physician contact was most acceptable and effective. Early in the project, we contacted physician groups that represented those segments of the physician population most likely to be affected by the review tools, if our methods were widely implemented. The groups included both the national level and state level primary care medical societies. Each group appointed a "liaison physician" to work with us in the development, implementation and evaluation work. We did not seek official endorsement from each group, but rather looked for thoughtful input that might represent the typical views of the members of each organization.

We recognized the importance of good communication throughout the course of the project. We held regular meetings with physicians in each state, first to introduce the project, then later to elicit their feedback about the claims and records criteria, and finally to present preliminary results. We also established a newsletter, the *DEMPAQ Quarterly*, to communicate regular updates to not only the physician groups, but to other interested organizations.

We discovered that the old adage "all politics is local" applied to this project. The state level liaison physicians identified issues, concerns and medical practice styles that were unique to each state. Nonetheless, we were able to achieve consensus and agreement, at the local level, with a set of review criteria based on national practice guidelines and policies. Sources of the guidelines included specialty society publications, medical journal articles and medical texts.

Performance Measures

How to develop measures

We developed performance measures that used two existing data sources to measure the quality of care: Medicare claims and physician office records. By using currently available data sources, review began without the delays imposed by new data collection efforts.

Both claims and records should be used to measure performance because they complement one another. Neither source alone is adequate. Medicare claims provide a broad view of medical care received by patients from all the providers involved in their care, including all visits, tests and procedures for all patients within a state. Relative to office record information, claims data are inexpensive to acquire although their analysis requires high level computer skills. Much of what the Medicare claims lack—clinical detail and information about drug prescriptions—the medical record provides. The record gives an in-depth view of a patient's care under the auspices of a single physician or group of physicians. The medical record includes information on prescription drugs, test results, physician reasoning and patient reactions to physician advice. While the record gives a close-up view of medical care, these data are more expensive to collect than claims data, requiring copying of the physician's office records.

What to measure

The DEMPAQ tools measure the technical quality of care. This approach assesses whether physicians made appropriate clinical decisions for patients and then correctly implemented the decisions. Because of our focus on technical quality, the DEMPAQ performance measures are clinically detailed. Each measure applies to a selected set of patients who share a specific clinical condition.

Our measures are also oriented towards processes of care known to, or strongly believed to, produce better patient outcomes. We reasoned as follows. Physicians will not know how to improve unsatisfactory outcomes unless the processes known to improve these outcomes have already been identified through prior research. We can provide better direction for quality improvement by measuring performance for those processes shown to produce better outcomes than by measuring the outcomes themselves.

We derive our criteria for measuring the quality of processes from practice guidelines or policies. The criteria are explicit. Each criterion is applied to the care received by each patient so that the patient is classified as receiving, or not receiving, care conforming to the guidelines. These data are summarized for patients receiving care from an individual physician to create a physician level performance rate; patients who receive care conforming to the criterion are counted in the numerator of the rate, and all patients eligible for this criterion, whether or not they received care conforming to the criterion, are counted in the denominator of the rate. The individual physician performance rates are then averaged to create a group performance rate. Thus our measures can be used to assess average conformance to practice guidelines.

There are sound reasons for our focus on average performance. First, it is often difficult to distinguish whether the performance of one physician is significantly different from other physicians because each physician typically has few cases in any given patient category. Second, greater gains can be made more efficiently by "moving the mean" of group performance. Third, identifying outliers is an expensive and often counterproductive exercise. Physicians see such efforts as intrusive and confrontational. Finally, working together, groups of physicians can achieve greater improvements by removing system barriers to good practice than by persuading individual physicians to try harder to overcome the faulty systems.

Whom and where to measure

Our performance measures evaluate the quality of care delivered by primary care physicians to Medicare care beneficiaries in the office setting. We designed our measures to provide clinically detailed information to physicians (or other providers of care) that would show where change is needed.

Our strategy does not require that the performance of every physician be measured. For the record review, average performance can be determined by measuring the quality of care delivered by a random sample of physicians who voluntarily agree to submit records in response to an invitation.

Based on the results of our project, a review of volunteer physicians (who were randomly invited to participate) will provide a representative view of the quality of care delivered to beneficiaries in a state. We initially used claims data to draw a random sample of physicians, stratified by specialty group, and a random sample of their Medicare patients. Approximately one-third of those physicians invited to participate chose to submit records. Subsequently we used claims data to compare the responders with non-responders. We discovered that participants represented their colleagues in many respects, and that the quality of care, based on claims, did not differ between responder and non-responder physicians.

Evaluating measures

Evaluation is a crucial and often overlooked step in performance measure development. We used the methods described below to evaluate and specify improvements in the performance measures.

We used four methods to evaluate the validity of claims-based measures:

- Critique of the clinical content by physicians liaisons;
- Critique of the data specifications by researchers with special expertise in using claims data;
- Test of agreement between the claims data and the medical record; and
- Structured physician review of sample cases, comparing the interpretation based on the claim to the interpretation based on the record.

We also used four methods to evaluate the records-based measures:

- Critique of clinical content by physician liaisons;
- Tests of inter-abstractor reliability both within and between the study states;
- Assessment of the adequacy of records documentation; and
- Structured physician review of a sample of cases judged to be non-conforming to the review criteria.

We recommend that these evaluations be repeated periodically to test the validity of measures over time as medicine and practice guidelines change.

Many of our claims and records review criteria demonstrated levels of reliability and validity that support their further use after only minor revisions. Examples include the measures related to care for chronic diseases, follow-up of key clinical events, prescribed drugs and preventive care. For follow-up of abnormal laboratory tests, the physician evaluations by structured review revealed the need for some specific revisions to the criteria.

The structured physician reviews (both claims and records) also supplied important information which aid the interpretation of performance measure results. This information quantifies the frequency with which patient, practice or other factors modify the interpretation of results.

Feedback Results of Performance Measurement

How to conduct feedback
As a result of the ongoing interaction and dialogue between project staff and physicians, many of the liaison physicians judged DEMPAQ to be a worthy, important project and chose to become involved in the feedback strategy. Acceptance by local physicians is critical to the success of any feedback strategy.

Successful feedback also depends on having at least one strong opinion leader physician as part of the process. In our experience this opinion leader plays a key role, particularly in a group feedback setting such as a CME workshop. He or she can set the tone for the workshop and remind participants of the purpose of the workshop, if necessary.

We tested a feedback strategy consisting of an integrated continuing medical education (CME) package with the following components:

- A CME workshop;
- A CME workbook; and
- Reports of profiles and patterns of performance based on Medicare claims data and office medical records.

As part of this workshop, physician participants worked through actual case records for their own or for anonymous patients, applying the DEMPAQ records tool. This exercise not only gave participants the opportunity to learn quality review by doing, but also gave us valuable feedback about how well the tool worked.

What information should be presented
Feedback of data on average performance is consistent with our philosophy of "moving the mean" of group performance rather than identifying physicians tagged as outliers. This strategy moves away from assigning blame towards a philosophy where every problem or error presents an opportunity for improvement. HCFA promotes this approach as part of the Health Care Quality Improvement Program (HCQIP).

We also presented clinically detailed information. The workshop, workbook and reports were oriented towards providing clinicians specific details about their performance. The records reports and office practice profiles gave the physicians their own data side-by-side with that of their peer group, a format which permitted quick comparisons. Workshop attendees reported finding this level of detail very helpful.

In general, we found the quality of care rendered in the ambulatory setting to be excellent. However, there were clearly areas where performance could be improved. These areas tended to be ones where medical practice was changing. Physicians constantly require information to stay abreast of the rapid changes in medicine.

Who receives the feedback?
The target audience for feedback was the primary care physician. We initially chose a CME framework for the feedback to encourage physicians to take part in the project. In return for their participation, physicians who submitted records were offered free CME credit. The DEMPAQ CME package was designed to meet both the needs of the project and the requirements of the Accreditation Council for Continuing Medical Education (ACCME).

Where to conduct the feedback?
We chose to present feedback in a CME setting. Our approach, basing the CME on the results of performance measurement, differs markedly from traditional approaches in which the selection of CME topics is influenced by the interests of researchers, or manufacturers of drugs or devices—rather than by the assessed needs of the average primary care physician.

Next steps
Future development and implementation of DEMPAQ will require five new activities:

- revise the performance instruments based on the results of reliability and validity testing;
- transfer the updated performance measurement technology to the PROs;
- transfer the technology to evaluate reliability and validity to the PROs;
- design and evaluate physician quality improvement campaigns; and
- design and evaluate consumer campaigns aimed at educating beneficiaries to be active contributors to the quality of care they receive.

The first step involves revising and updating the performance instruments based on the results of the reliability and validity studies performed during the project. Recommendations include changes to the criteria (both records and claims), the decision rules (records), the scoring (records), and coding (claims).

The updated tools and methods for performance measurement must then be transferred to the PRO program with appropriate manuals and support materials. We found the cost of conducting at least the record review to be reasonable. Assuming 20 patients per physician and 100 physicians per state, with measurements of average performance made every two years, the total cost per state for the record review is estimated to be between $70,000 and $94,000.

In addition, the tools for maintaining the data quality and validity need to be transferred. Without data quality checks, the results of measurement may be unusable. The measurement tools must also be continuously validated, as medical practice changes rapidly. Without ongoing validation, the tools quickly become outdated. National and specialty groups and state physicians societies should be involved in the process of continuous revision.

Finally, quality improvement campaigns involving both physicians and consumers should be developed, implemented and rigorously evaluated. Good health outcomes result from an effective partnership between physician and patient. To improve the quality of health, both physician and patient must adopt parallel approaches to improving quality.

In summary, we developed performance measures based on fundamental principles of quality improvement over the three-year DEMPAQ project. As the structure of the U.S. health care system changes over the next few years, this innovative work can form the foundation for continued efforts to improve quality of care.

Chapter 1
Conclusions and Recommendations

In 1989, the Health Standards and Quality Bureau (HSQB) encouraged Peer Review Organizations (PROs) to submit proposals for pilot projects to test methods to fulfill the mandate of Congress for physician office review. We formed a consortium of researchers and three PROs, and together proposed a project to *D*evelop and *E*valuate *M*ethods for *P*romoting *A*mbulatory Care *Q*uality (DEMPAQ). In this chapter we present 11 recommendations based on three years of work in developing, implementing and evaluating methods of performance measurement directed toward improvement of quality for the primary care given by general medical physicians to Medicare beneficiaries. These recommendations are listed here and discussed below.

- Implement a program of quality improvement, especially for care of chronic diseases.
- Measure performance of primary care.
- Measure performance of processes of care known to produce better patient outcomes.
- Involve physician groups in the development of performance measures.
- Focus on improvement of the average performance of physician groups.
- Use two existing data sources: claims and records.
- Use clinically meaningful and explicit performance measures.
- Rely upon volunteer physicians for records samples.
- Evaluate the performance measures with collaboration from physician groups.
- Give responsibility for review to external review organizations composed of locally practicing physicians in a state.
- Plan Implementation of quality improvement in primary care now.

In addition we discuss issues in implementing the DEMPAQ review strategy on a national level, and note how our project fits into health care reform.

Implement a Program of Quality Improvement for Primary Care

Review of quality in physician office practice for Medicare beneficiaries was mandated by Congress in the Omnibus Budget Reconciliation Act of 1986. Review of ambulatory care in Health Maintenance Organizations (HMOs) began in 1989. To date, however, Medicare has not addressed care received in physicians' offices, although improving the quality of care is no less important for beneficiaries receiving fee-for-service care than for beneficiaries receiving HMO care. In 1993, the Health and Human Services (HHS) Office of the Inspector General issued a report strongly recommending that review of ambulatory care begin.

We recommend that HCFA implement a program of quality improvement focused on chronic diseases common among Medicare beneficiaries for which prevention and treatment are effective in improving patient outcomes. The program should be directed toward helping primary care physicians to keep up-to-date with medical advances, and using measures of conformance to practice guidelines. Such a program fits naturally into HSQB's new Health Care Quality Improvement Initiative (HCQII).

In Chapter 2 we give an overview and rationale for the performance measures we developed and tested.

Measure Performance of Primary Care

We recommend a focus on primary care, specifically, the care given by general medical physicians, for three reasons.

Majority of Ambulatory Care Visits are for Primary Care

Over 49% of the ambulatory visits for the Medicare population are to primary care physicians, i.e., general and family practitioners and internists (U.S. Department of Health and Human Services, 1992). These visits contribute substantially to the annual costs of ambulatory care in the U.S., which now exceed the annual costs of inpatient care. We used the national statistics on frequency of utilization by the Medicare population to target our review toward common diseases and commonly used drugs, tests and procedures.

Opportunity for Prevention, Early Detection and Intervention
Decisions made in primary care, typically the point of first contact with the health care system for a patient, can save lives and limit disability through prevention and early intervention in disease. This not only benefits patients but also averts costly secondary and tertiary care. For instance, a randomized controlled trial reported by the Diabetes Control and Complications Trial Research Group in 1993 demonstrated that tight control of glucose metabolism through intensive general medical care prevents blindness and renal failure due to diabetic microvascular disease (The Diabetes Control and Complications Trial Research Group, 1993).

Opportunity to Help Office-based Physicians Keep Up with Advances in Medicine
Because most physician training takes place in the hospital setting and prepares physicians for careers as specialists, many primary care physicians enter practice with too little relevant preparation. Those whose practice is conducted primarily in their private offices are more isolated from pertinent continuing medical education or from peer oversight than their hospital-based colleagues. They find it difficult to keep up with rapid changes in medical technologies in all the fields covered by specialists. Therefore, we developed measures of physician conformance to clinical practice guidelines and other frequently updated sources of clinical guidance in the design of a review strategy that helps primary care physicians to keep current with advances in primary care practice.

Measure Performance of Processes of Care Known to Produce Better Patient Outcomes
To reach the goal of quality improvement, we chose to measure performance for processes of care known to produce better patient outcomes. We reasoned as follows. If unsatisfactory outcomes are reported to physicians, they will not know what action to take to improve care unless processes known to improve these outcomes have already been found through prior research. We can provide better direction for quality improvement by measuring performance for those processes shown to produce better outcomes than by measuring the outcomes themselves.

Our choice to focus on processes of care was supported when physician groups in each state reviewed our reports on performance and identified areas in the average level of performance that justified efforts to improve the process of care.

Involve Physician Groups in the Development of Performance Measures
We worked with physicians from state and national medical societies and societies of internal and family medicine to design performance measures that were credible, supported by local physician leaders, and clinically meaningful to physicians in each state. We conclude that this strategy worked (see Chapter 3). Physician volunteers from relevant societies in each state and at the national level collaborated actively in the project. Moreover, many of them were sufficiently convinced of the clinical relevance of the work so they further collaborated in the small trial of use of the DEMPAQ performance measures in continuing medical education workshops. (This trial was the subject of a separate project report due in December 1993.)

In Chapter 3 we discuss how we organized our collaboration with physician groups.

Focus on Improvement of the Average Performance of Physician Groups
We recommend that the goal of review be improvement of the average performance of primary care physicians in a state. This should be achieved by collaborating with physician groups to promote continuous improvement within their peer group through traditional collegial means.

We designed performance rates because we were interested in average performance rather than isolated instances of poor performance. By aggregating individual performance rates to create average performance rates for groups of physicians, we targeted areas where the average physician can improve. We chose this strategy for both philosophical and practical reasons.

Statistical Limitations
It is often difficult to distinguish whether the performance of one physician is significantly different from other physicians because each physician typically has few cases in any given patient category.

Suitability of Performance Measures for Quality Improvement Activities
We propose that quality improvement activities be at the level of physician groups. Thus measurement of group performance is needed to identify areas where average performance could be improved.

Expense of Intervention to Improve Quality
It is costly to intervene for each of many specific cases of poor performance, or for many specific physicians with apparently poor performance. Moreover the overall gain in quality per intervention is small. Furthermore, intervention by external review organizations with individual physicians is resisted by physicians who see it as intrusive and confrontational.

Physicians' Responsiveness to Peers and Peer Group Leaders
It is likely to be more effective and less expensive to interact with physician leaders and rely on them to lead the group to change behavior since physicians are known to respond to the example of their peers.

Removal of Barriers to Conformance
Barriers to conformance with good practice often are embedded in systems for delivering care. Working together, groups of physicians can achieve greater improvements by removing system barriers to good practice than by persuading individual physicians to try harder to overcome the faulty systems.

Use Two Existing Data Sources: Claims and Records
We recommend using two existing data sources: Medicare claims data and the records currently kept by physicians in their offices. By using currently available data sources, review can begin without the delays imposed by new data collection efforts.

Both claims and records should be used because they complement one another. Neither source alone is adequate. Claims data document the occurrence of all visits, tests and procedures for all patients within a state. Records data generally report only the care given by one physician to a panel of patients, however they provide a more comprehensive and clinically meaningful picture since they may include information on prescription drugs, test results, physician reasoning and patient reactions to physician advice.

Even when claims and records data appear to measure the same item of performance, the records measure may be more precise in identifying those patients for whom care can be improved. For example, in measuring performance of Pap smears using claims data, we can compute the percentage of women for whom a claim for a Pap smear was submitted of those women with a visit in a given year who fall within a relevant age group. However, when intervening to improve quality, it is useful to separate performance rates for those women who refused a Pap smear. Patient refusal cannot be identified in the claims data for a given year, while it is likely to be retrievable from the medical records data.

Claims data should be used for a continuing overview of relevant areas of practice to identify areas for quality improvement, and to evaluate if quality improvement activities are followed by improvement in performance. Samples of records data should be used for periodic estimation of average performance in greater clinical detail in order to estimate the magnitude of measurement error in the claims-based measures. Records data need also to be used to address the wider set of clinical issues that can be tracked in medical records data. Appropriate samples of records data should be identified using the claims data.

Use Clinically Meaningful and Explicit Performance Measures

The approach demonstrated in the DEMPAQ project differs from other common approaches to external review of quality of care in three important ways.

Clinically Meaningful Performance Measures

DEMPAQ performance measures are different from physician profiles of crude rates of utilization of service. Each DEMPAQ measure applies to a selected set of patients who share a specific clinical condition. A criterion for quality, derived from practice guidelines or policies, is applied to the care received by each patient so that the patient is classified as receiving, or not receiving, care conforming to the guideline. These data are summarized for patients receiving care from an individual physician to create a conformance rate; patients who received care conforming to the criterion are counted in the numerator of the rate, and all patients eligible for this criterion, whether or not they received care conforming to the criterion, are counted in the denominator of the rate. The individual physician conformance rates are then averaged to create a group conformance rate. Thus the DEMPAQ project can measure conformance to practice guidelines.

Explicit Criteria to Evaluate Quality of Care for Both Claims and Records Review.

We used claims data to create statewide profiles of average physician conformance with tests, visits and procedures recommended by clinical practice guidelines for patients with specific conditions. We also produced profiles of performance by specialty groups. (By merging with zip code data, similar profiles could be produced by region within each state.)

We developed an interactive computerized system to abstract relevant data from photocopies of medical records covering two years of visits and to apply explicit criteria for conformance to practice guidelines. The review system included appropriate use of 38 classes of drugs, response to abnormal results for 13 diagnostic tests, appropriate diagnosis and management of six chronic diseases, periodic performance of 4 items of preventive care, appropriate performance of two office procedures and follow-up of important clinical events.

Comparison of Physicians' Office Practice Profile with Case Mix Adjustment

Many efforts to compare physicians' practices for use of services neither refer to differences in patient characteristics, nor adjust for case mix. Instead we compare practice profiles after adjusting for case mix using the Ambulatory Care Groups, and using resource value units to sum the cost implications of the use of combinations of services (see Chapter 4).

In Chapter 4 we describe the DEMPAQ Claims-based Review system (DCRS), in Chapter 5 we discuss the related methodological issues, in Chapters 6–8 we give the claims-based results for each of the three states and in Chapter 19 we present an analysis of patterns of variation in claims-based performance measures.

In Chapter 11 we describe the DEMPAQ Records-based Review System (DRRS) and related methodological issues, in Chapters 16–18 we present the records-based results for each of the three states, and in Chapter 20 we present an analysis of patterns of variation in the records-based performance measures.

Rely on Volunteer Physicians for Records Samples

We used the claims data to draw a random sample of physicians, stratified by specialty group and geographic region within a state, and a random sample of their Medicare patients. These physicians were asked to participate voluntarily in the project by submitting records of the selected patients for review. About one-third of those invited chose to participate. We used data from the state's provider file, and from our claims-based measures to determine whether these physicians are representative of their colleagues. As reported in Chapter 12, they are representative in many respects, including quality of performance on claims-based measures; i.e., there were no differences between responders and nonresponders.

The voluntary nature of the physician sample is an asset because those physicians who were willing to volunteer their records in order to review their own performance are more likely to appreciate the usefulness of the measures and also more likely to become leaders of quality improvement efforts among their colleagues.

We do not recommend attempting to sample care of every physician in rotation. If the goal of review is to detect problems with average performance rather than to identify individuals who perform much worse than their colleagues, then it does not matter that some physicians are never reviewed and others are reviewed several times. This approach is analogous to industrial quality control methods where samples of a product are periodically tested to detect proper functioning of the production process.

In theory, one would prefer to review a representative sample of physicians, including those who will and will not voluntarily submit records. However, the benefits of opportunities for improvement in patient outcomes found through working with volunteer physicians outweighs the shortcomings of a skewed sample.

We saw an example of this principle during the continuing medical education workshops designed to provide feedback to the physicians who volunteered to submit records on performance. At the workshop, physicians are guided through review of their own records in order to critique the validity of the performance measures. Having had the opportunity to appraise for themselves the validity of the measures, and their usefulness as a guide to opportunities for quality improvement, these physicians become potential allies and leaders when quality improvement activities are promoted within their peer group. (This modification to the DEMPAQ project was reported in December 1993.)

In Chapter 12 we report the comparison of the physicians who responded to our invitation to submit records for review to those who were invited but refused.

Evaluate the Performance Measures with Collaboration from Physician Groups

Systems to measure quality of care must be carefully evaluated to determine whether the information they provide is useful and accurate. We recommend that the methods described below be used to evaluate and identify specific improvements in performance measures. These evaluations should be repeated periodically to test the continuing validity of the measures over time, as practice guidelines are updated and as review personnel may change.

Validity of the Claims-based Measures

We used four methods to evaluate the validity of claims-based measures.

- Physician liaisons from state and national physician societies critiqued the clinical content of the measures.
- Selected researchers with expertise in using claims data critiqued the data specifications.
- We tested agreement between medical records and claims data to determine the reliability of claims data for performance measurement.
- We tested the sensitivity and specificity of the measures as a test for performance problems by conducting structured physician review of samples of cases that were identified as conforming and not conforming to the claims review criteria.

Evaluation of Records-based Measures

We also used four methods to evaluate the records-based measures.

- Physician liaisons from national and state medical societies critiqued the draft measures for clinical validity.

- We conducted tests of inter-abstractor reliability both within and between the study states for abstraction from records of the data used for performance measurement.
- We tested the specificity of the measures as a test for performance problems by conducting structured physician review of samples of cases that were identified as not conforming to the review criteria.
- We included, as a component of the record review system, an assessment of the adequacy of documentation in each record for five types of information. Analyses of the patterns of records documentation adequacy indicate areas where medical records should be improved.

Many of the claims and record review criteria demonstrated levels of reliability and validity that support their further use after only minor revisions, for example the measures related to care for chronic diseases, follow-up of key clinical events, prescribed drugs and preventive care. For follow-up of abnormal laboratory tests, the physician evaluations by structured review revealed deficiencies but also revealed whether and how these deficiencies could be lessened by specific revisions of the criteria. We also developed a record review module for prompt and appropriate response to potentially serious presenting complaints, including cough, chest pain, dizziness, difficulty breathing, fatigue, and joint pain. This function was set aside during the early phases of record review as not possible to complete within the project resource limits. Nevertheless, the complaints module relates to the risk of missing serious causes of common complaints. Specific revisions could make this a valuable addition to office review. Specific recommendations for refinements of the measures are listed in Chapter 11 and in technical appendices to this report.

The results of our evaluations of claims-based measures are given in Chapters 9 and 10, and results of the evaluations of records-based measures in Chapters 13 through 15.

Give Responsibility for Review to External Review Organizations Composed of Locally Practicing Physicians in a State

PROs are well-suited to conduct clinically meaningful explicit measurement to promote improvement in quality of average physician performance. We pioneered the testing of a review strategy in which PROs collaborated with local medical societies. Our project demonstrates the successful operation of this model, which is now official policy of the HSQB as part of its Health Care Quality Improvement Initiative (now known as the Health Care Quality Improvement Program [HCQIP]).

PROs can be defined generally as external quality review organizations formed by locally practicing physicians. We conclude that both features of the PRO are useful for promotion of quality. As external review organizations, PROs are required to be advocates for the health care consumer. However, PROs are generally owned by and employ locally practicing physicians who understand the local clinical environment and can act as opinion leaders for their colleagues. PROs are now required to develop long-term collaborative relationships with the local physicians whose care is being reviewed. PROs in the Fourth Scope of Work will also employ persons with the technical skills needed to collect and analyze data, and to conduct collaborative quality improvement activities with medical societies.

Plan Implementation of Quality Improvement in Primary Care Now

While planning for national implementation of quality improvement in primary care could begin immediately based on our work, some aspects of performance measurement require further development, including revisions based on physician structured review of the measures.

Data Adequate for Many Review Items.

We found the existing data sources to be adequate for many review items but not for others. Furthermore, our evaluation of methods reveal some specific improvements in the two data sources that could improve feasibility and validity of review.

During the course of the DEMPAQ project, the Health Care Financing Administration (HCFA) has continued to improve the Medicare claims databases. We recommend that these efforts continue and be directed specifically to improving the usefulness of these data for purposes of quality review. We also recommend careful consideration of means to improve the quality of physician record keeping to support delivery of high quality of care to Medicare beneficiaries. We note that the documentation needed to support good care, in an era when many physicians are likely to share in the care of a patient, is the same as the documentation needed for purposes of quality review.

Specific details are given in Chapter 5 for claims and Chapter 11 for records.

Further Programming Needed for Widespread Implementation

For claims-based measures, further programming work is needed to make the procedures sufficiently user friendly for analyses to be conducted by individual PROs.

For records-based measures, the interactive data system was taught during the project to nurse reviewers in the three study PROs. Thus it has already been shown to be feasible for review to be conducted by individual PROs. However, as noted above, the physician evaluation of the measures identified that some of the measures need revision and refinement. Furthermore, the analysis of the records data files to apply the review criteria in the form of computer algorithms, and to produce reports of record review results in tabular and graphic formats was conducted by research staff. Creating a seamless and user friendly mechanism that would enable a PRO to do this work will require further development including programming.

Development of Methods for Intervention to Improve Quality

Actual intervention to improve quality was not within the scope of this project which was limited to the development and testing of methods of measurement suitable for use in continuous quality improvement. Pilot testing of intervention strategies will be needed.

Issues in Implementing the DEMPAQ Review Strategy on a National Level

Planning national implementation of the DEMPAQ review strategy should take into account the following issues.

Refinement of the Performance Measures

As noted above, researchers will need to develop user friendly software that would enable the average PRO to operate this review technology. Dr. Weiner, whose expert computer staff conducted the current claims-based review, estimates that this could require approximately 18 months of work and approximately $500,000 in direct costs.

For the records-based review, the physician evaluations of measures showed that minor revisions to most indicators are needed; for other measures more substantial work would be required. Dr. Lawthers, who managed the building, dissemination and evaluation of the interactive computerized record review system, estimates that the large set of measures relating to chronic diseases, prevention, drugs, and follow-up could be ready for widespread use with minor revisions after nine months of work and $100,000 in direct costs. The module for follow-up of abnormal test results needs more substantive, specific revisions and pilot testing. These activities including evaluation would take nine months and $100,000 in direct costs. The module for review of response to presenting complaints could be revised and pilot tested with final evaluation within one year with $100,000 in direct costs. Creating a user-friendly system for producing records review reports could take six months of effort and $70,000 in direct costs.

Training of Personnel

Initial training of PROs to take over the performance measurements could probably be accomplished in a national workshop taking several days for each of the claims and records review measures. We recommend that the early phases of individual PRO use of the measures be accompanied by close oversight and evaluation of success in operating the technology. As mentioned above, ongoing testing of the reliability and validity of the measures is required. PROs would also need some specific training in working collaboratively with local medical societies on promoting quality in primary care. This could likely be provided in a two-day workshop.

We recommend that HCFA consider a stepwise introduction of quality improvement for primary care, with groups of PROs starting review in sequence, so that lessons learned by the first PROs can be used to benefit PROs initiating this activity at a later date.

Operating Costs

Based on our analysis of the costs of operating the DEMPAQ review system, we found that the total marginal cost of review (additional costs beyond the PRO's general operating costs) once the review system was developed and disseminated to the Alabama and Iowa PROs, ranged from $35 to $47 per review. Moreover, the review time varied widely based on the complexity of the patient's record. Therefore we recommend using records-based measures to review only a sample of physicians each year.

Chapter 21 presents our analysis of the costs of implementing the DCRS and DRRS.

Sample Sizes for Records Review

Based on the variability in physician performance reflected in the record review results, we could compute the physician sample size needed to detect a five-point improvement in average performance. This could be done for each of the many records-based measures. For example, in the DEMPAQ project we asked participating physicians to submit records for 25 patients that we had selected. If we assume we receive 20 records per physician and 95 percent confidence limits, the number of physicians that we would need to sample for monitoring of a newly prescribed drug for seven classes of drug related measures varies from 25–157. For monitoring of a preexisting diagnosis (six diagnoses reviewed) under the same assumptions, the number of physicians needed ranges from 10–37.

Further details of the variability of physician performance on different measures and the resulting sample sizes needed under various assumptions are reported in Chapter 20.

Improved Data Sources

While better claims and records data could improve the validity and feasibility of reviewing the quality of care in physicians' offices, we conclude that useful review can begin with existing data sources. Moreover, the Medicare claims data have been improved considerably since we began our project.

Acceptability to Physicians

The DEMPAQ project clearly demonstrates that state and national medical societies will collaborate with PROs in quality promotion activities that relate to clinically useful measures of average physician performance. Medical societies can convince themselves of the clinical usefulness of measures if they are involved in their development, evaluation and use. Since PROs themselves are organizations comprised of locally practicing physicians, such collaboration is both logical and useful. In the DEMPAQ project, medical societies were often deeply mistrustful of the PRO program initially. In particular, they expressed serious concerns that quality measures would be used to sanction, or alter the reimbursement for, individual physicians. In the end, the project's collaborative relationships with medical societies were established through the leadership of physicians who already worked as peer reviewers for the PRO.

There are no grounds for believing that this review strategy would work if conducted by nonphysician organizations or by organizations concerned with physician reimbursement or with application of sanctions.

Applications to Other Settings and in Health Care Reform

While the specific design of health care reform will only be decided after a long legislative process, the measurement of quality will play an important role, and primary and preventive care will be a priority. We developed the DEMPAQ measures to promote quality in primary care delivered in physician offices to Medicare beneficiaries. However, the generic format of the measures makes it relatively easy to adapt them to other groups of physicians, patient populations and medical conditions. Since the measures focus on average performance, they can be immediately applied to sum up the performance of HMOs and health care plans. As such, they are ideally suited for inclusion in "report cards" that will be used by the proposed health alliances, employers and consumers to choose among health care plans. The DEMPAQ strategy for ambulatory care is perfectly suited for use by the proposed regional professional foundations that will provide a voluntary approach to quality improvement.

The legislation proposed by the President provides for continuation of fee-for-service practice, and for Medicare and PRO review to continue initially. We have demonstrated a method of performance measurement suited for quality improvement in this situation; we believe it fits HCFA's needs now.

Chapter 2
Overview and Philosophy of the DEMPAQ Project

Abstract
The Omnibus Reconciliation Act of 1986 established that Peer Review Organizations would review care provided in physicians' offices. In response to this congressional mandate, the Health Care Financing Administration funded several pilot projects to test the viability of alternative approaches to review of care rendered in the physician's office. DEMPAQ is one of those pilot projects.

This chapter discusses the importance of ambulatory care quality. DEMPAQ methodology moves us away from the traditional review for quality improvement of identifying single instances of poor care. Instead, DEMPAQ provides rates of performance from medical record and claims data so that physicians can monitor their own performance, singly and as a group.

DEMPAQ, a project to *D*evelop and *E*valuate *M*ethods for *P*romoting *A*mbulatory *C*are *Q*uality, is a three-year collaborative demonstration project, emphasizing education and feedback to physicians, concentrating on the quality of care rendered in office settings. DEMPAQ was begun on September 1, 1990, and concludes its first phase with the submission of this report. A modification to the project on feedback and education to test one intervention strategy was completed by December 15, 1993.

The DEMPAQ contract was awarded to the Delmarva Foundation for Medical Care, Inc., the federally designated PRO for the state of Maryland. Collaborating PROs on the project include the Iowa Foundation for Medical Care and the Alabama Quality Assurance Foundation. Academic researchers responsible for designing and carrying out the work with the PROs include R. Heather Palmer, M.B., B.Ch., S.M., Harvard School of Public Health, Principal Investigator; Jinnet Fowles, Ph.D., from the Park Nicollet Medical Foundation; Deborah Garnick, Sc.D., from the Brandeis University; Ann Lawthers, Sc.D., from the Harvard School of Public Health; and Jonathan Weiner, Dr.P.H., from the Johns Hopkins School of Public Health.

Importance of Quality

Why measure health care quality? Measuring the quality of health care services has never been more important. As insurers, employers, and the state and national governments address the problems of cost of, and access to, care, an increasingly vocal segment of the population raises the issue of quality. Will it be possible to reduce the rate of increase in the cost of health care without adversely affecting the quality of care?

It is important to consider advances in medical technology when dealing with quality of care issues. Physicians face the challenge that the practice of medicine is continually changing. As new technologies and treatments revolutionize medicine, they must learn when and how to use them appropriately. New technologies are often expensive and can be potentially dangerous; using them appropriately and implementing them safely is an issue of quality. Clinging to outdated, ineffective and sometimes risky treatment in the face of newer, safer alternatives is also an issue of quality. Physicians need a way to monitor whether their own practices are keeping pace with medical advances. DEMPAQ's goal was to develop and evaluate performance measures that could be used for that purpose.

Since the inception of the Medicare program in 1967, billions of dollars have been spent in the name of measuring and ensuring the quality of health care services (Lohr, 1990). Much of this expenditure has been devoted to looking for quality of care problems in the hospital setting. Until now, federal policymakers, hospital administrators, accrediting bodies and academic researchers have emphasized hospital care for a variety of reasons. First, the hospital focus offers several obvious poor outcomes (such as mortality or iatrogenic infection). Second, it is relatively easier to collect data from approximately 5,300 hospitals than from 530,000 physicians (U.S. Bureau of the Census, 1992). Third, the 1970s and early 1980s saw the largest increase and fastest growth in health care expenditures in the hospital setting (U.S. Bureau of the Census). Finally, until recently, there has been a lack of recognition that pre- or post-hospital care strongly influences hospital use and outcomes.

The Importance of Ambulatory Care

Is review of ambulatory care quality important? We believe it is. The success or failure of care rendered in the ambulatory health care setting significantly determines future care and health care costs.

First, ambulatory care provides opportunities for prevention and early detection which may improve patient outcomes while saving the costs of treatment for more advanced stages of disease (Palmer, 1988).

Second, the vast majority of individuals receive medical care in a physician's office rather than in a hospital. In 1990, approximately 75% of Americans contacted a physician on an ambulatory basis while only 10% experienced a hospital admission (USDHHS, 1991).

Third, poor clinical management in the ambulatory setting can result in higher health care costs in the future, often with a worse outcome for the patient. For example, if the primary care physician does not correctly manage congestive heart failure, the patient could require an expensive hospitalization. If diabetic complications affecting the retina of the eye are not recognized and treated promptly, some patients will lose their eyesight. Even less serious errors in clinical care cause patients needless limitation of health, anxiety, waste of personal time and dissatisfaction.

Fourth, poor coordination of ambulatory care creates wasteful redundancy of clinical services, as when a specialist repeats expensive tests without realizing that they were already performed by the patient's primary care physician.

Fifth, the percentage of all resources expended in the ambulatory care environment is growing annually and approaching 50% of all health care resource consumption (Sonnefeld, 1991). Since the implementation of Prospective Payment System (PPS), care and related resources have been shifted from hospital to ambulatory care settings.

Sixth, problems of underutilization and poor access to care can be identified by assessing ambulatory care.

Finally, malpractice suits can result from missed ambulatory diagnosis or inadequate ambulatory treatment. The costs of malpractice include not only the award to the patient, but legal and insurance fees.

In view of its importance, it is surprising that ambulatory care has only recently begun to be subjected to quality review. The potential for quality improvement is great. Recent research provides evidence of rectifiable deficiencies in ambulatory care even in well-regarded large group practices, hospital outpatient units and HMOs (Palmer 1988, McDonald 1976, Barnett 1978). The extent of deficiencies among physicians less exposed to collegial contacts is likely to be greater. Many office based physicians are in one or two person offices and generalists in particular may only have infrequent contacts with colleagues in hospitals.

The Importance of Rates of Performance

The traditional strategy for quality improvement has been to identify single instances of poor care and to target remedial programs at the individual. This strategy of finding "bad apples" seeks to assign blame for performance deficiencies in individual cases to individual physicians (Berwick, 1989). The "apple lopping" strategy is an expensive and labor-intensive technique that undermines the positive relationship needed for continuous quality improvement. It is for these very reasons that the PRO Fourth Scope of Work for hospital review has moved away from individual case findings to patterning of care in the new Health Care Quality Improvement Program (HCQIP). In this initiative, improvement will occur by PROs working cooperatively with the medical, hospital, and beneficiary communities to focus on areas for quality improvement.

Using rates to measure quality avoids problems associated with traditional review. Rates describe how often an event occurs in a specific population, e.g., the proportion of diabetic patients receiving a HgbA1c test in the prior 12 months. When the numerator of a rate reflects how often a physician performs a recommended clinical activity, such as ordering a test, and the denominator is the number of patients who are eligible for this activity, the resulting rate may be called a measure of the quality of performance.

To be reliable and valid, performance rates should be constructed from explicit criteria that applied to specific cases. Explicit criteria lead to consistent reviews over time and between reviewers. They are visible and can be challenged, discussed, altered and improved by those involved in the development process, thereby ensuring at least the face validity of the measures. In addition, the actions to be taken to improve quality are readily seen.

Rates prove particularly useful in quality measurement because they permit a variety of comparisons. Rates may be compared over time, between providers or institutions, or compared to a predefined standard. Further more, rates can be constructed to reflect either desirable or undesirable processes or outcomes.

DEMPAQ Philosophy

For the DEMPAQ project, we built the performance measures on three basic principles:

- ❖ The content of performance measures should be clinically useful and focused on the quality of health care in relation to specific patient characteristics rather than on crude utilization rates;
- ❖ The process for developing the measures should be interactive and open, using a "customer-supplier" dialogue between user and developer; and
- ❖ The measure should be designed for education and feedback for all physicians rather than for identifying and sanctioning outliers.

Quality Not Utilization

Many traditional measures of performance are focused on utilization, e.g., hospitalization rates or case-mix adjusted surgery rates. In contrast, DEMPAQ performance measures are focused on the technical quality of medical care in relation to the specific characteristics of individual patients. This approach assesses whether a physician made an appropriate clinical decision for a patient and then correctly implemented the decision, i.e., doing the right thing ***and*** doing it right. The focus on quality is designed to identify practices where a change in physician action, i.e., change in the process of care, could lead to improvement in outcomes.

Because of the focus on technical quality, the content of the DEMPAQ performance measures is clinically detailed and oriented toward processes of care. The processes we selected for review are relevant to the everyday practice of medicine in the ambulatory setting. This emphasis is crucial if the performance measures are to be useful to practicing physicians.

Open Black Boxes

Most commercial review systems are "black boxes" where the process for aggregating the review data takes place inside a computer with the actual computer algorithm hidden from view. In contrast to the black box approach, DEMPAQ has developed an open system, one in which both the criteria and the processes for scoring are readily observable. The cornerstone of the development process is a "customer-supplier" dialogue between user and developer. As will be noted in Chapter 3 of this report, the evolution of the DEMPAQ performance measures has been marked by the active involvement of many physician groups.

Educational Feedback

DEMPAQ reinforces the strong commitment among physicians for professional self-examination and self-improvement. It provides feedback about patterns of performance averaged over many instances of care, many patients, and many physicians. It enables physicians, either as groups or individuals to compare their performance to that of their peers, and to track whether they have improved their performance over time. Feedback of performance comparisons fosters the ambition to do as well as, or better than, one's peers.

The primary intention of DEMPAQ's measures is to convince physicians of the opportunities for improvement by promoting professional self examination and assisting in a campaign for improvement through educational workshops. Rather than focusing on the performance of individual physicians, DEMPAQ measures concentrate on moving the mean of group performance.

This strategy has advantages. First, there is movement away from assigning blame and toward implementing one of the tenets of the total quality management philosophy—every problem or error presents an opportunity for improvement (Berwick, 1989). Second, the strategy of focusing on patterns of practice rather than individual errors is also being promoted by HCFA as part of the Health Care Quality Improvement Program. This statewide educational strategy has significant potential to cause improvement in the quality of care for large number of patients.

What is DEMPAQ's Structure?

Two sources of information are used in the DEMPAQ review process: Medicare claims and physician office medical records of Medicare patients. These data sources are complementary; each has advantages and disadvantages. Medicare claims provide a broad view of the medical care received by patients. Claims cover care in all settings: in-hospital, ambulatory, skilled, nursing facilities, surgi-centers, rehabilitation centers and others. Visits to office-based physicians, hospital outpatient department services, laboratory tests, imaging tests and physical therapy treatments are among the items in the ambulatory data. Relative to medical record information, claims data are inexpensive to acquire. For each service, information is included on the date of service, who provided it, location (office, hospital, etc.), charges and patients' diagnoses.

Much of what the Medicare claims lack—clinical detail and information about drug prescriptions—the medical record provides. The medical record gives an in-depth or magnified view of a patient's care under the auspices of a single physician or group of physicians. Physician impressions and thought processes are part of the medical record. While the record gives a close-up view of medical care, these data are more expensive to collect than claims data and require intrusion in a physician's office for the information.

The physician forms the unit of analysis for both the claims and records review for the DEMPAQ project. For example, we might measure the rate by which physicians appropriately follow-up the prescription of drug X. We construct this rate for each physician we are studying and then aggregate the data across groups of physicians. The task of constructing the database which permits analysis at the physician level is approached differently for claims and records.

Claims Approach

Creating the claims performance measures requires data manipulation, since the raw Medicare data are stored at the level of the claims transaction. A single claim may contain many services for a particular patient. For example, it may contain charges for a visit to a physician as well as charges for laboratory or imaging tests. These services must be disentangled and a file constructed that contains a single line for each unique service. A patient's claim will have many lines, each capturing different services, such as an office visit, a lab test or physical therapy. Claims also contain administrative data that are used for billing purposes only. These so called "adjudication" data must be identified and purged from the database.

Claims are filed on behalf of patients, or beneficiaries. To develop physician-oriented analyses or "profiles" for DEMPAQ, patients are assigned to the provider who delivers the majority of a patient's primary care. This strategy ensures that physician performance rates will not include the care of patient they see only incidentally. Because of the project's focus on primary care, DEMPAQ identified physicians in the following categories: internal medicine; family practice; general practice; and osteopathic medicine. When two or more primary care physicians provided an identical number of visits to a patient, the amount of ancillary services provided was used to break the tie, thus identifying the physician known as the patient's "primary care source." Overall, approximately 80% of the primary care visits were provided by the patients' primary care source.

In addition to Medicare Part A and Medicare Part B claims data, DEMPAQ used data from Medicare's computerized beneficiary file and from carrier provider files. The beneficiary file contained information about the patient, such as date of birth. The provider files give demographic information about the physician such as specialty and the uniform provider identification number (UPIN). A full description of the DEMPAQ Claims Review System (DCRS) is provided in Chapter 4.

Medical Records Approach

The medical records data are collected and organized differently than the claims information. A random sample of physicians, selected from a listing of all "primary care source" physicians, was invited to participate. The letter of invitation was accompanied by a randomly sampled listing of 25 of the physician's Medicare patients. If the physician chose to participate in the DEMPAQ project, she/he photocopied (or had the PRO photocopy) notes for a two-year period (July 1, 1989, to June 30, 1991) from the records of the randomly sampled patients. Once the records were received at the PRO, the data were abstracted using a computerized record review system, described further in Chapter 11.

Profile Development

Claims

In developing profiles from the claims data, we base DEMPAQ performance measures on clinical events, specific actions and activities that have clinical significance for physicians. We chose this strategy so that the results of the review may aid in identifying targets for physician education.

A key consideration in developing performance measures from claims is the nature of the claim. For example, because Medicare does not cover prescription drugs, DEMPAQ could not develop performance measures involving drugs.

We group the claims performance measures into three types of profiles: condition specific; office practice; and preventive care services. Chapters 6, 7 and 8 provide claims profile books which reflect family practice claims profiles for each of the three study states. These profiles were provided to each physician who participated in the project.

Medical Records

As with the claims performance measures, the performance measures derived from the office medical record were selected for their clinical relevance. DEMPAQ constructed the measures at three different levels:

1. *Function*—activities that physicians typically perform in the course of an office visit, such as making a diagnosis, responding to an abnormal test, prescribing and monitoring a drug;

2. *Clinical Items*—specific examples of the functions; for instance, hypertension, diabetes mellitus and congestive heart failure are clinical items within the diagnosis function; and

3. *Indicators*—the clinical processes that influence patient outcomes. Each indicator is a rate of performance for a process of care; for example, two indicators for the Drug Function are:
   - ❖ Drug dose and frequency appear in the record; and
   - ❖ Drug is prescribed for an appropriate condition.

Unique criteria have been developed for the review of each clinical item/indicator combination. The criteria answer the question "What represents appropriate performance for this indicator and for this item?" For example, the actions which represent appropriate monitoring for a diagnosis of diabetes differ from the actions appropriate in monitoring ischemic heart disease.

Throughout the remaining chapters in this report, the reader will be provided detailed descriptions of, and analyses from, the DEMPAQ Record Review System (DRRS) and the DEMPAQ Claims Review System (DCRS). Analyses of the results of those review systems will provide information on the agreement between the medical record and claims; patterns of variation for claims and records and cost of implementing DCRS and DRRS. In the following chapter, we will discuss the collaborative approach which was taken for this project.

Chapter 3
Building a Collaborative Project

Abstract
The federal Health Standards and Quality Bureau (HSQB) has a legislative mandate to review, through the Peer Review Organization (PRO) program, the care given to Medicare beneficiaries in physicians' offices. To address the mandate, a collaborative project was proposed, in which a "customer-supplier" dialogue would shape the ultimate form of the review methods.

This chapter describes the nature of the collaborative relationship developed over the course of the DEMPAQ project and presents the various strategies used by the project to maintain communication.

The federal Health Standards and Quality Bureau (HSQB) has a legislative mandate to review, through the Peer Review Organization (PRO) program, the care given to Medicare beneficiaries in physicians' offices. Although there is a substantial body of research and prior experience on methods of review in ambulatory care, even in physicians' offices (Palmer, 1988), it is not known what difficulties will emerge in applying this research and experience to the particular circumstances of PRO review of office-based care in the 1990s. For these and other reasons, HSQB began a strategy of pilot testing methods for office-based review. One of these projects is DEMPAQ, a project to develop, implement, and evaluate review methodologies that could be used for mandatory office-based review of care received by Medicare beneficiaries. The findings of the project could also be used to derive recommendations for full-scale implementation of such a review system within the PRO program.

To accomplish the goal of developing a useful, feasible review system, the project was required to address the needs of several different constituencies, including the PRO program and the physician community. In addition, the PRO program needed to call upon the resources of the research community to aid in the design and implementation of the project. Thus a collaborative project was proposed, one where a "customer-supplier" dialogue would shape the ultimate form of the review methods. Physicians were the customers of the review methods; while the PROs and researchers, as developers and testers of the methods, were the suppliers. PROs, at times, acted as customers for the technology developed by the researchers. The following describes the role of each of these groups in the development and testing of the DEMPAQ methods.

History

The project was originally proposed to HSQB on March 7, 1990, with subsequent revisions on July 6 and 19, 1990. Initially the project included a total of ten PROs working with health care researchers under the direction of Principal Investigator, R. Heather Palmer, M.B., B.Ch., S.M. The ten PROs met on a number of occasions and provided input to Delmarva Foundation for Medical Care, Inc. (DFMC), during preparation of the proposal. A significant source of support and assistance during this time was the American Medical Review Research Center (AMRRC), whose mission includes fostering collaborative research between the PRO and research communities. In many ways AMRRC facilitated the PROs' joining hands with the researchers.

Budget considerations during the negotiations eventually led to the funding of only three PROs for the project. HCFA recommended that the Alabama Quality Assurance Foundation and the Iowa Foundation for Medical Care be included, along with DFMC as the prime contractor. The remaining seven PROs became an Advisory Committee which met periodically during the project. Under this structure, DEMPAQ officially began on September 1, 1990. The PRO community has remained extremely interested in this project and is eager to learn the DEMPAQ review systems as well as when additional PRO implementation of physician office review will occur.

Customers: Physicians

Physician input has been critical to the project. A major goal of DEMPAQ has been to convince the physician community of the need for self-improvement. Thus the project has developed a variety of measures to assess physician performance. To be truly useful to physicians, measures need to be clinically relevant and actionable, that is, they must address aspects of performance that are important to and under the control of the physician. Furthermore, the results of the performance measurement needed to be credible to physicians if the project were to succeed in convincing physicians of the need for self-improvement. For these reasons and because the project's performance measures would ultimately be used by physicians, the project sought physician input during all phases of the project. Perhaps the most important initial task of the researchers and PROs was to engage the physician community in the DEMPAQ project at the national level and state levels in Maryland, Alabama and Iowa.

Developing Working Relationships.

Soon after the project began in September 1990, DFMC hosted a meeting in Washington, D.C., to announce the project and invite national, state and county medical societies. Dr. Palmer, Linda E. Clark, CEO of DFMC, and Dr. Christian Jensen, DFMC's Corporate Medical Director, made presentations about the project. The presentations focused on the research nature of the project and that participation would be voluntary. Medical societies were encouraged to become involved in the project and assured that their participation was critical to its success. While much skepticism was voiced, especially because of the PROs' involvement in the project, the presentations alerted medical societies and physicians to DEMPAQ aims, in particular, that DEMPAQ genuinely sought input.

From September 1990 through February 1991, multiple presentations were made across the country to national organizations and to the states' specialty societies. The meetings were coordinated through the PROs with presentations made primarily by Dr. Palmer. As a physician and a researcher, she was successful in helping physicians recognize that the DEMPAQ project provided a unique opportunity for the physician community to work with the federal government and health care researchers to design a nonintrusive office review of Medicare patients. Through repeated assurances by Dr. Palmer and the PROs, the project was ultimately able to attract key specialty and medical society participation. These included:

- *National*
  American Medical Association (AMA)
  American Society of Internal Medicine (ASIM)
  American Academy of Family Physicians (AAFP)
  American Osteopathic Association (AOA)
  American College of Physicians (ACP)

- *Maryland*
  Medical-Chirurgical Faculty of Maryland (MedChi)
  Maryland Academy of Family Physicians (MAFP)
  Maryland Society of Internal Medicine (MSIM)

- *Alabama*
  Medical Association of the State of Alabama (MASA)
  Alabama Chapter of the American Academy of Family Physicians (AAFP)
  Alabama Chapter of the American Society of Internal Medicine (AlaSIM)

- *Iowa*
  Iowa Medical Society (IMS)
  Iowa Clinical Society (ICS)
  Iowa Academy of Family Physicians (IAFP)
  Iowa Osteopathic Medical Association (IOMA)

All these groups participated in the project with the understanding that their participation in no way constituted an endorsement of the project.

In developing DEMPAQ physician resources, three specific levels of physician contact were identified and used throughout the project.

Liaison Physicians

Liaisons with the AMA, ASIM, AAFP, AOA and ACP at the national and state levels. These physicians served as active participants during all three years of the project.

Participant Physicians

This group included physicians who responded affirmatively to the invitation to join the project and who submitted medical records.

Physician Reviewers

Physicians recommended by their primary care specialty societies to assist in various evaluation functions of DEMPAQ.

The role of these resources is described more completely below.

Physician Feedback

Feedback from physicians began at the first meetings held in each of the three study states where physicians voiced concern about the project's goals and agenda. This feedback continued during the design, implementation and feedback phases of the project.

Formal feedback was solicited when the initial profile specifications and record review criteria were shared with liaison members of the medical and data communities. Liaisons reviewed and critiqued the performance measures and offered suggestions for their revision. The version of the criteria used by the DEMPAQ Record Review System (DRRS) and by the DEMPAQ Claims Review system (DCRS) claims profiles stemmed from this feedback process.

During project implementation, physician reviewers in each state (see above for definition) provided peer review of cases flagged by DRRS and assisted in the evaluation of the DCRS performance measures included in the profiles. These functions and the results of the feedback are fully described in Chapters 10 and 14.

Feedback was also solicited at special liaison physician briefing sessions held in the fall and winter of 1992-93 (see Chapter 15). Further feedback was sought during the Feedback and Education Contract Modification to the DEMPAQ Project. The report of these activities was submitted in December 1993.

Communicating with Physicians

Timely communication with physicians was deemed essential to the goals of the project and several means of communication developed. Early in the project, a "Question and Answers" paper was drafted to respond to the most frequently asked questions about the project. This document was also used as an enclosure to the letter sent to the 900 physicians who were invited to participate in the project (see Appendix 1).

A second source of communication was the *DEMPAQ Quarterly*, a quarterly newsletter widely distributed to over 500 individuals or organizations. These newsletters chronicled each phase of the project and served to provide current information to the medical, legislative, PRO and research communities. These newsletters began in November 1990 and continued throughout the project (see Appendix 2).

A third source of communication was periodic news stories in specialty society newsletters and journals.

Suppliers: PROs and Researchers

Successful implementation of a project the size and scope of DEMPAQ required intensive collaboration between researchers and PROs. While the investigators developed the review tools and evaluation methodology, the PRO implemented the record review, worked with state liaison physicians, physician reviewers and made local presentations.

The investigative research team was carefully assembled by Dr. Palmer. Each investigator brought a special area of expertise to the project. Regular communication between the investigators was facilitated by weekly conference calls and periodic meetings in Boston and Baltimore. Communication between the investigators and the PROs was facilitated by written reports and contact sheets.

The process of developing the research tools and evaluation strategy involved close contact between the PROs, particularly DFMC, and the researchers. As implementation proceeded first in Maryland, and then in Iowa and Alabama, daily phone conversations between the review coordinator at Delmarva and the research director in Boston were the norm. This meant that the progress of the project was constantly monitored by PROs and researchers.

Summary

As a collaborative venture, DEMPAQ has worked remarkably well. Initially reluctant and suspicious physicians have become helpful, if somewhat cautious, collaborators. Our experience with this project demonstrates several points:

- Physicians, if kept well informed and offered ample opportunities to become involved in the process, are willing to work collaboratively with PROs and researchers on issues of quality.
- A PRO/researcher collaboration is not only productive but educational as PROs learn research tools and methods and researchers learn "real world" concerns and issues.
- Open and frequent communication with all parties is essential to the success of a project.

Appendix 1

«DATA mdlst2»

DELMARVA
FOUNDATION
FOR MEDICAL CARE, INC.

9240 CENTREVILLE ROAD   EASTON, MD 21601   PHONE: (410) 822-0697   FAX: ADM. (410) 822-7971 FAX: REVIEW (410) 822-1997

July 26, 1991

«Name»
«Street»
«City», «State»   «ZipCode»

Dear Doctor,

As President and Medical Director of the Delmarva Foundation for Medical Care, Inc., I am writing to invite you to participate in a project to develop and evaluate methods to promote ambulatory care quality. This project is becoming nationally known as DEMPAQ, and concerns the quality of general medical care delivered to Medicare beneficiaries in physicians' offices. It is being conducted under the joint leadership of the Harvard School of Public Health and the Delmarva Foundation for Medical Care, which is the Peer Review Organization for Maryland. A number of additional researchers from Johns Hopkins University and the Park Nicollet Foundation are assisting in the project. Peer Review Organizations in Alabama and Iowa will be field testing the review instruments next year. The Medical and Chirurgical Faculty of the State of Maryland, the Maryland Academy of Family Physicians, and the Maryland Society of Internal Medicine are knowledgeable concerning this project and have appointed liaison physicians to the project.

You were randomly selected for this project from a list of physicians in Maryland who submitted Medicare claims for care provided between July 1, 1989, and June 30, 1991. We ask that you begin your participation by submitting copies of the medical records for the enclosed list of your patients for the period of July 1, 1989, through June 30, 1991. You will be paid $5 per record. The attached instructions should assist your staff in responding to this request. As the project proceeds, you will receive feedback on your practice as evaluated from these records, compared to that of your colleagues in Maryland. You would then serve as the peer reviewer for these cases. The results of your review will be used to redesign the project's instruments. If you complete this review and an associated education module, you will receive four hours of category 1 CME credit at no cost to you.

The results of this study will not be used to judge your practice nor to deny payment. It will merely be used to assist in the development of a review methodology which may be used by other peer review organizations in the future. This study will not be used to compare costs of medical care among providers.

All information submitted will be strictly confidential. Data linked to you personally will not be released to HCFA or anyone outside the research group.

I welcome you to the project and encourage your active participation. Please contact me if you have any questions about the project, the attached instructions, or the actual listing of records being requested.

Sincerely,

C. E. Jensen, M.D., M.P.H.
Corporate Medical Director

BALTIMORE
REGIONAL
OFFICE
7467 RIDGE ROAD
SUITE 130
HANOVER, MD 21076
PHONE: (410) 684-3362
FAX: (410) 684-2078

WASHINGTON, DC
REGIONAL
OFFICE
1000 THOMAS JEFFERSON ST., NW
SUITE 507
WASHINGTON, DC 20007
PHONE: (202) 337-2375
FAX: (202) 337-3768

DELMARVA FOUNDATION
FOR MEDICAL CARE, INC.

INSTRUCTIONS FOR DEMPAQ PARTICIPATION

Participation in the DEMPAQ project means that we ask you to submit photocopies of a sample of your Medicare medical records. Just as your name was selected at random from a list of primary care physicians who treated Medicare patients, the patient records have also been selected randomly. The attached list shows the first six letters of the patient's last name and his/her Medicare number. If the list does not contain enough identifying information, please contact us at the telephone number listed below.

To reduce the copying burden, we are prepared to either reimburse you for copying costs at $5 per record, or, if you desire, arrange to have a DFMC representative come to your office to copy the records at no cost to you. If you copy the records, please complete and submit the enclosed invoice. If your costs exceed this reimbursement rate, for items such as excessive postage, please let us know.

We would like to stress that any and all information submitted is strictly confidential. While a report of this trial will be sent to the Health Standards and Quality Bureau in the Health Care Financing Administration in July 1993, it will contain only aggregate results. Individual data will be released only to single participating physicians.

We ask that you submit one copy of the medical record for each of the patients shown on the attached list for the time period of July 1, 1989, through June 30, 1991. To ensure a complete and accurate review of your patterns of practice, we request that you include data such as:

- visit notes,
- problem lists,
- laboratory and radiology reports (July 1, 1989, through June 30, 1991),
- hospital discharge summaries (July 1, 1989, through June 30, 1991),
- consult reports (July 1, 1989, through June 30, 1991), and
- medication summaries (July 1, 1989, through June 30, 1991).

Please mail your records to:

**Delmarva Foundation for Medical Care, Inc.
9240 Centreville Road
Easton, Maryland 21601**

Attention: Gail Karge, Medical Records Supervisor

**For further information regarding the project,
please contact Nellaine Price, RN, at 1–800–999–3362.**

July 26, 1991

DELMARVA FOUNDATION
FOR MEDICAL CARE, INC.

DESCRIPTION OF THE DEMPAQ PROJECT

What is DEMPAQ?

DEMPAQ stands for Development and Evaluation of Methods to Promote Ambulatory Care Quality. It is a three year collaborative demonstration project funded by the Health Care Financing Administration (HCFA) to develop tools to review physician office care. The project emphasizes education and feedback to physicians.

The academic collaborators responsible for designing and carrying out the work include renowned researchers from Harvard (R. Heather Palmer, M.D., Principal Investigator, and Deborah Garnick, Sc.D.), Johns Hopkins (Jonathan Weiner, Dr.P.H.) and Park Nicollet Medical Foundation (Jinnet Fowles, Ph.D.).

The PRO collaborators include the Delmarva Foundation for Medical Care, Inc. (DFMC) of Easton, Maryland, the prime contractor, the Iowa Foundation for Medical Care, and the Alabama Quality Assurance Foundation.

Physician Consultants have been invited from national and local physician groups. Additionally, collaborators are being sought from the physicians who participate in the review process. Doctors will be asked to provide input concerning the design of the review instruments and will be critiquing the results of the review process for relevance, validity and review burden.

Project Design: Claims Review

DEMPAQ consists of two review methods: claims profiling and medical record review. Medicare Part B claims data will be used to generate computerized profiles of quality and utilization indicators. The profile format will be submitted to liaison members of the medical community for critique and comment prior to dissemination of the data to the profiled physicians. Carrier groups will be asked for input on technical details.

The profiles will then be fed back to physicians invited to participate in the record review, along with comparison information for peers. Physicians will be asked to review and critique the profiles for relevance and usefulness relative to quality of care.

A small number of physicians in Maryland will be asked to submit a sample of their medical records so that we may explore the accuracy of the claims profiling technique as a quality indicator. This process will involve peer review of the claims profile and the record.

Project Design: Record Review

The medical record review focuses on giving feedback to physicians comparing them to their peers for the adequacy of record documentation and performance of routine clinical functions such as testing, prescribing and monitoring drugs, procedures and diagnosis. A computer will be used by each reviewer to abstract information from the photocopied medical record.

As with the claims review, the record review criteria will be submitted to members of the medical community for review and comment prior to implementing the actual record review. This process is well underway. Following the record review, the data collected will be fed back to the reviewed providers and their comments and critiques solicited.

Who will be reviewed?

DEMPAQ plans to develop methods to review the care delivered to Medicare patients by primary care physicians with an office practice. Specialists, whose care is already subject to intensive review in the hospital, are explicitly excluded. General Practitioners, Family Practitioners and Internists with a practice in General Medicine are most likely to be eligible for review. Of course, doctors without Medicare patients will not be reviewed.

Is participation mandatory?

Participation in the record review is voluntary. Physicians contacted for the record review will be selected randomly. For each of these physicians, a random sample of twenty-five patient records will be selected. Doctors will be offered reimbursement for each photocopied record to defray the costs of processing and mailing. Additionally, project staff will be available to help with the copying effort for those physicians for whom the burden of copying records is excessive. All information submitted will be strictly confidential. No physician or patient specific information will be released by the project staff, except to the physician in question for educational purposes.

How will the review data be used?

Since DEMPAQ's purpose is educational, results comparing an individual's performance to that of his or her peers, will be fed back confidentially to that physician only. Aggregate data will be included in the final report describing the progress of the project.

Physicians groups may wish to incorporate the findings of the project into educational programs and perhaps produce educational materials relating to issues that emerge from the claims and record review. For example, they may wish to use the information generated by the DEMPAQ project regarding practice patterns in the development of practice guidelines. Data that spring from the project may include information such as the typical treatment profile for elderly hypertensives. This information may be of interest to physicians choosing among a variety of options in caring for the elderly patient. Or, perhaps medical groups will wish to make recommendations regarding appropriate levels of record documentation based on the results of the record assessment.

The possible uses of the data generated by the study for improving the quality of medical care are many and exciting.

July 26, 1991

Appendix 2

DEMPAQ

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 1, Number 1 November, 1990

WHAT IS DEMPAQ?

DEMPAQ stands for Development and Evaluation of Methods to Promote Ambulatory Care Quality. It is a three year collaborative demonstration project funded by the Health Care Financing Administration (HCFA) to develop tools to review physician office care. The project emphasizes education and feedback to physicians.

The academic collaborators responsible for designing and carrying out the work include renowned researchers from Harvard (R. Heather Palmer, M.D., Principal Investigator, and Deborah Garnick, Sc.D.), Johns Hopkins (Jonathan Weiner, Dr.P.H.) and Park Nicollet Medical Foundation (Jinnet Fowles, Ph.D.).
The PRO collaborators include the Delmarva Foundation for Medical Care, Inc. (DFMC) of Easton, Maryland, the prime contractor, the Iowa Foundation for Medical Care (IFMC), and the Alabama Quality Assurance Foundation.

Physician collaboration is invited from national and local physician groups. Additionally, collaboration will be sought from the physicians who participate in the review process. Doctors will be asked to provide input concerning the design of the review instruments and will be critiquing the results of the review process for relevance, validity and review burden.

PROJECT DESIGN: CLAIMS REVIEW

DEMPAQ consists of two review methods: claims profiling and medical record review. Medicare Part B claims data from the Common Working File will be used to generate computerized profiles of quality and utilization indicators. The profile format will be submitted to liaison members of the medical community for critique and comment prior to sharing the data with the individual profiled physicians. Carrier groups will be asked for input on technical details. Physicians will receive their own profiles along with comparison information for peers. Physicians will be asked to review and critique the profiles for relevance and usefulness relative to quality of care.

A small number of physicians in Maryland will be asked to submit a sample of their medical records so that we may explore the accuracy of the claims profiling technique as a quality indicator. This process will involve peer review of the claims profile comparing it to the record.

PROJECT DESIGN: RECORD REVIEW

The medical record review focuses on giving feedback to physicians comparing them to their peers on the adequacy of record documentation and performance of routine clinical functions such as testing, prescribing and monitoring drugs, procedures and diagnosis. A laptop computer will be used by each reviewer to abstract information from photocopied medical records.

For example, the computer program might prompt with the following types of questions to evaluate a chart with the prescription of a drug to a Medicare patient. Is the current dose and duration documented? Is the drug appropriate to diagnoses and other drugs? Are there any contraindicated drugs or disease states?

DEMPAQ IS A REVIEW OF

- Care for Medicare beneficiaries
- Office-based practices
- General medical care
- Quality and utilization

Linda E. Clark, RN, Executive Vice President
Delmarva Foundation for Medical Care, Inc.

R. Heather Palmer, M.B., B.Ch., S.M.,(British equivalent of MD), Director
Center for Quality of Care Research and Education, HSPH

As with the claims review, the record review criteria will be submitted to members of the medical community for review and comment prior to carrying out the actual record review. Following the record review, the data collected will be fed back to the reviewed providers and their comments and critiques solicited.

WHO WILL BE REVIEWED?

DEMPAQ plans to develop methods to review the care delivered to Medicare patients by primary care physicians with an office practice. Specialists, whose care is already subject to intensive review in the hospital, are explicitly excluded. General Practitioners, Family Practitioners, and Internists with a practice in General Medicine are most likely to be eligible for review. Of course, doctors without Medicare patients will not be reviewed.

The physician groups listed above will be contacted for their input concerning various phases of the work.

IS PARTICIPATION MANDATORY?

Individual physicians will be asked to participate in the record review process by submitting photocopies of patient records. In return, they will receive feedback about their practice patterns and level of record documentation as compared with their peers. They will be asked to comment on and critique the usefulness and relevance of the information. For this level of participation we are exploring the possibility of offering category I CME credit.

Participation in the record review is voluntary. Physicians contacted for the record review will be selected randomly. For each of these physicians, a random sample of twenty patient records will be selected. Doctors will be offered reimbursement for each photocopied record to defray the costs of processing and mailing. Additionally, project staff will be available to help with the copying effort for those physicians for whom the burden of copying records is a problem. All information submitted will be strictly confidential. No physician or patient specific information will be released by the project staff, except to the physician in question for educational purposes.

Claims profiles will be mailed to eligible physicians with Medicare patients for their personal information.

HOW WILL THE REVIEW DATA BE USED?

Since DEMPAQ's purpose is educational, results comparing an individual's performance to that of his or her peers, will be fed back confidentially to that physician only. Aggregate data will be included in the final report describing the progress of the project.

Physician groups may wish to incorporate the findings of the project into educational programs and perhaps produce educational materials relating to issues that emerge from the claims and record review. For example, they may wish to use the information generated by the DEMPAQ project regarding practice patterns in the development of practice guidelines. Data that spring from the project may include information such as the typical treatment profile for elderly patients with hypertension. This information may be of interest to physicians choosing among a variety of options in caring for the elderly patient. Or, perhaps medical groups will wish to make recommendations regarding appropriate levels of record documentation based on the results of the record assessment.

The possible uses of the data generated by the study for improving the quality of medical care are many and exciting.

HOW DOES DEMPAQ RELATE TO CPR?

DEMPAQ is a completely separate endeavor from the Comparative Performance Reports (CPR) to be released by HCFA. CPR involves payment issues while DEMPAQ's agenda is quality. CPR focuses on "outlier" physicians while DEMPAQ feeds back information to the whole spectrum of primary care physicians.

KEY ELEMENTS OF DEMPAQ

- Review is educational
- Review tools developed with national and local physician inupt
- Results of both claims and record review to be critiqued by reviewed physicians; critiques will aid redesign of tools
- Claims review unobtrusive
- Record review based on random selection of physician - participation encouraged

DEMPAQ TIMELINE OF ACTIVITIES AFFECTING PHYSICIANS

PROJECT COMPONENT	ESTIMATED START* MARYLAND	ESTIMATED START* IOWA/ALABAMA
NEWSLETTER		
Quarterly newsletter to medical groups	November 1990	December 1990
CLAIMS PROFILING		
Proposed claims profiles to medical groups for review and critique	October 1991	October 1991
Aggregate claims profiles to medical groups	April 1992	August 1992
Individual claims profiles to physicians	May 1992	September 1992
Receive physician responses	September 1992	November 1992
VALIDATION OF CLAIMS PROFILES		
Request medical records to link to claims data	January 1992	N/A
RECORD REVIEW		
Record review criteria to medical groups for review and critique	February 1991	February 1991
Request medical records for review	May 1991	September 1991
Aggregate results of record review to medical groups	April 1992	August 1992
Individual results to physicians	May 1992	September 1992
Receive physician responses	August 1992	November 1992

Actual Start Dates May Vary

STAY TUNED....

Future editions of the DEMPAQ newsletter will include updates on the progress of the project.

IF I HAVE QUESTIONS WHO DO I CALL?

If you have any questions about the research design or how the project might affect you, please call Ann G. Lawthers, Sc.D., Research Director, QCaRE, Harvard School of Public Health, 617-432-1205.

Questions about the project implementation should be directed to Jean E. Edwards, Project Manager, Delmarva Foundation, 800-999-3362.

DEMPAQ QUARTERLY

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 1, Number 2 March 1991

DEMPAQ is a three-year collaborative project to develop methods to review the care given to Medicare beneficiaries by physicians with office-based practices. The project focuses on education and feedback to physicians. Two review methods are being developed — medical record review and claims profiling.

RECORD REVIEW UPDATE

The *DEMPAQ Review System* (DRS) consists of two related tools for reviewing office-based medical records: the Record Documentation Assessment (RDA) and the Clinical Performance Assessment (CPA).

Record Documentation Assessment

The RDA will examine the office-based medical record for a set of basic information items that are well-accepted parts of all medical records. For quality and continuity of care it is essential that physicians keep records which maintain key clinical facts about a patient. In addition, physicians who have been sued find the level and excellence of medical record documentation critical to a successful defense.

Record Documentation Assessment — Criteria

We propose six criteria for record documentation (see below). The criteria shown are not new for DEMPAQ but appear in recommendations of other physician groups such as the draft , "Principles for Documentation of Ambulatory Encounters," proposed by the Ambulatory Review Committee of the American Medical Peer Review Association. We believe a review system such as DRS will be useful to physicians by giving them feedback on how their medical records compare to those of their colleagues.

RECORD DOCUMENTATION ASSESSMENT CRITERIA

Are the following in the record?

- **Date of visit**
- **Patient complaint or reason for visit**
- **Observations regarding the patient**
- **Conclusions or diagnosis**
- **Plan for patient care or evaluation**
- **Identification of clinician**

Record Documentation Assessment — Scoring

The goal of the RDA is to report to individual physicians their average performance on record documentation across all their patients. We plan to score the RDA on a continuous scale, e.g., 0 to 3, where 0 would equal no documentation and 3 would equal the highest levels of relevant documentation. The final scoring scale will depend on input from physician organizations. The method proposed for calculating RDA scores would mean that a single patient's record is not given undue weight, while consistently poor patterns of documentation across all patients would be portrayed.

Clinical Performance Assessment

The second component of DRS is the Clinical Performance Assessment. To understand the CPA, one must first understand some of the terminology used to describe the CPA.

Clinical Performance Assessment — Functions

CLINICAL FUNCTIONS TO BE REVIEWED BY CPA

- **Prescribing Drugs**
- **Ordering Tests**
- **Following Management Plans**
- **Performing Procedures**
- **Working Up Patient Complaints**
- **Making Diagnoses**
- **Providing Prevention/Screening**

The Clinical Performance Assessment (CPA) is based on the fact that patient encounters are built around a number of things that physicians do: so-called functions (see above). For this purpose, a function is defined as a broad category of clinical activities typically performed by the doctor in the

Record Review Update, con't on page 2

Linda E. Clark, RN, Executive Vice President
Delmarva Foundation for Medical Care, Inc.

R. Heather Palmer, M.B., B.Ch., S.M. (*British equivalent of MD*), Director
Center for Quality of Care Research and Education, HSPH

Record Review Update, con't from page 1

EXAMPLES OF PROPOSED CLINICAL ITEMS

Function	Clinical Item
Complaints	Dizziness Chest Pain Joint Pain
Diagnoses	Hypertension Ischemic Heart Disease Arthritis
Drugs	Diuretics Coronary Vasodilators Diabetes Therapy
Prevention/Screening	Influenza Vaccination Breast Exam
Procedures	Sigmoidoscopy Colonoscopy
Tests	EKG Serum Creatinine/BUN Serum Potassium

course of a patient visit. The occurrence of any of these seven functions will be identified and assessed by reviewers.

Clinical Performance Assessment — Clinical Items

The CPA checks each visit in the record for examples of each function, that is, "clinical items." Clinical items are defined as specific instances of a function. For example, some of the clinical items proposed for the complaint function are weakness, dizziness, and chest pain; for tests — EKG, serum potassium, serum creatinine/BUN; for diagnoses, hypertension, diabetes mellitus, and arthritis. These represent some of the most commonly occurring drugs, tests, procedures, complaints, and diagnoses found in the medical records of Medicare patients. We will be reviewing the list of clinical items with the medical community. As time and resources permit, additional clinical items will be added to the Clinical Performance Assessment.

Clinical Performance Assessment — Criteria

Draft criteria for each of the clinical items within a function are being developed in consultation with clinicians and by reviewing available relevant practice guidelines. As they are drafted, specific review criteria will be circulated to physician organizations for comment and critique, before they are confirmed for field testing.

Clinical Performance Assessment — Scoring

Each instance of a clinical item performed for a function is scored. The scoring scale proposed includes the categories Met, Acceptable Alternative, Not Reviewable, and Flagged for physician review. Scores are then summarized across cases to produce a measure of the adequacy of overall performance on that function, i.e., a rate of performance.

Clinical Performance Assessment — Performance Indicators

For each of the seven CPA functions, DEMPAQ will produce rates for acceptable performance of the function, called "indicators." For example, an indicator for drugs might be "drug prescribed for appropriate condition."

The choice of indicators is designed to provide useful feedback to the full range of physicians who give general

Record Review Update, con't on page 3

Clinical Performance Assessment (CPA) Diagnosis Function Summary — Hypothetical Example of Feedback

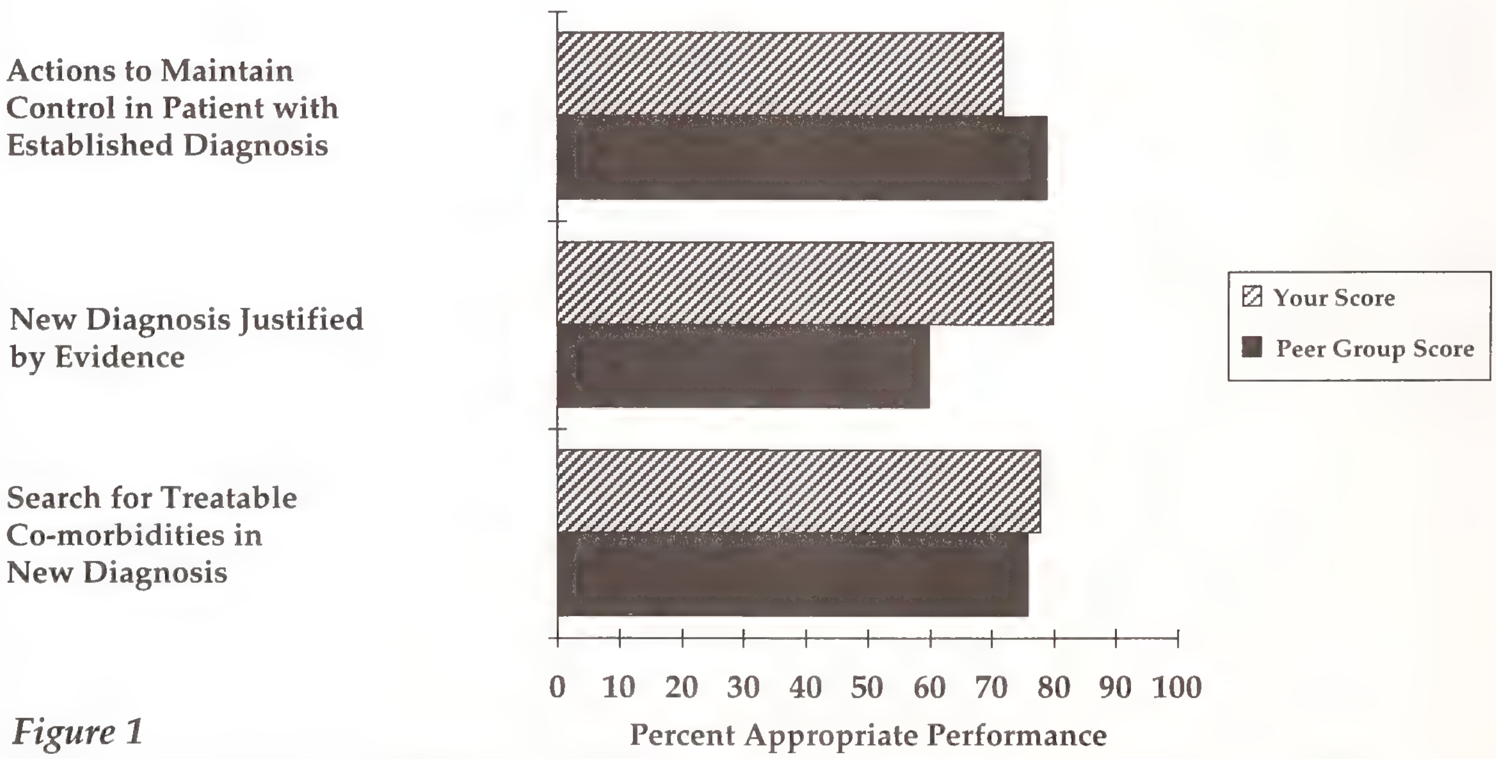


Figure 1

Record Review Update, con't from page 2

medical care. The emphasis is more on the implementation of decisions than on the correctness of decision-making. Figure 1 shows a hypothetical example of how an individual physician might receive feedback on performance as scored by the CPA.

Clinical Performance Assessment — How it Works

The following hypothetical record entry illustrates how the CPA might work.

> *Patient frustrated because of increased weight but she is not smoking. Taking BP meds as directed, following diet. Cream helping hands although not using as directed. Has woken up incontinent of urine x2, drinking tea at night. BP 140/95 — slightly improved. Will consider changing meds if incontinence continues. Return in four months.*

The review would proceed as follows.

- Since no new ***drugs*** are prescribed at the time of the visit, the review proceeds directly to the test function review. Blood pressure is one of the ***tests*** evaluated; review of the physician's response to the results of the elevated BP (review compliance and nutrition) shows appropriate performance.
- The visit note did not include a hospital discharge summary or a consultant's report: ***follow-up*** would not be evaluated.
- No ***procedures*** are performed so the review would proceed directly to evaluating the patient's complaints.
- At present the ***complaint*** of incontinence is not among the clinical items reviewed by DRS and would not be evaluated.
- The physician does not make a new ***diagnosis*** at this visit nor does he/she provide ***preventive care***; these items would not be scored.

Feedback

Rates of performance on both the RDA and CPA will be reported to providers along with the rates for their peers. This educational component is critical to the philosophy of the project. We will use the opportunity of this feedback to solicit comment from providers on the usefulness and value of the review. These comments will assist in the evaluation and redesign of the DEMPAQ Review System.◊

UPDATE: CLAIMS PROFILING

DEMPAQ plans two types of claims profiling: generic profiles and focused, condition-specific profiles. In order to establish a physician profile, Medicare beneficiaries will be associated with the primary care physician who provides most of the office care for the patient. The usefulness of this approach will be evaluated.

Generic Profiling

For each physician we plan to indicate the total number of patients seen and the number of visits. Generic measures may include rates of: hospitalization, visits to that physician, visits to other physicians, use of ancillary services, types of visits. These generic measures are utilization measures, but will be "adjusted" for characteristics of patient mix (clinical differences and socio-economic factors).

Physicians will be compared with other physicians in their specialty (e.g., family practice, internal medicine). When feasible we will also make these comparisons according to their urban/rural location of the practice.

These comparisons will give physicians a broad view of how their practice patterns compare to those of their colleagues.

Diagnosis-Specific Profiles

We are considering several focused measures. For example, for a particular diagnosis, we may produce overall visit rates and rates of specific activities (such as tests). We may produce rates of utilization of newer versus older tests with similar uses. These comparisons provide many more opportunities for clinically useful comparisons of practice patterns.

The two types of profiles being developed are for quality assessment and bear no relationship to the profiles developed under HCFA's Comparative Performance Reports which relate to physician payment.◊

***DEMPAQ Quarterly* is published on a regular basis to promote greater understanding of the DEMPAQ project.**

Comments or questions should be addressed to:

Ann G. Lawthers, Sc.D, Research Director
QCaRE, Harvard School of Public Health
677 Huntington Avenue
Boston, Massachusetts 02115

or

Jean E. Edwards, Project Manager,
Associate Executive Vice President
Delmarva Foundation for Medical Care, Inc.
341 B. North Aurora St.
Easton, Maryland 21601

DEMPAQ

QUARTERLY

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 1, Number 3 June 1991

DEMPAQ is a three-year collaborative project to develop methods to review the care given to Medicare beneficiaries by physicians with office-based practices. The project focuses on education and feedback to physicians.

RECORD REVIEW CRITERIA

As described in the last newsletter, the DEMPAQ record review process involves reviewing seven common functions typically performed by the physician in the course of treating the patient. For each of the seven Functions, DEMPAQ has proposed a series of quality indicators. These indicators do not measure quality directly but rather are a screen for activities which contribute to overall quality.

Indicator Development

We have sought help in developing the indicators from members of national and state medical organizations as well as project staff. To date, representatives from the American Medical Association, the American Academy of Family Physicians, the American Society of Internal Medicine, the American Osteopathic Association, and the American College of Physicians, have agreed to give input on the specifics of the record review criteria. In addition, individuals from the Medical and Chirurgical Faculty of Maryland, Maryland Academy of Family Physicians, Maryland Society of Internal Medicine, the Iowa Medical Society, Iowa Osteopathic Medical Association, Iowa Clinical Society, Iowa Academy of Family Physicians, Medical Association of Alabama, Alabama Society of Medicine, and the Family Practice Association of Alabama are DEMPAQ Physician Consultants. Advice from Consultants does not constitute official endorsement of the criteria by any of the parent organizations.

The indicators proposed for the DEMPAQ project are presented below. The selection of indicators is intended to provide useful feedback to physicians who give general medical care. Until we receive all feedback from all Physician Consultants they must be considered in draft form.

Indicators

Complaint Function
- Referral is made in specified time frame if required by nature of complaint
- Complaint is resolved appropriately in specified time

Diagnosis Function
- Actions are taken to monitor/maintain control in patient with old diagnosis
- New diagnosis is justified by evidence in medical record
- Actions are taken to search for treatable causes/co-morbidities of new diagnosis

Record Review Criteria, con't on page 3

Linda E. Clark, R.N., Executive Vice President
Delmarva Foundation for Medical Care, Inc.

R. Heather Palmer, M.B., B.Ch., S.M. (*British equivalent of MD*), Director
Center for Quality of Care Research and Education, HSPH

RECORD REVIEW TIMELINES

This summer we will begin inviting a random sample of physicians from Maryland to participate in the DEMPAQ project. Iowa and Alabama physicians will be contacted in the fall.

DEMPAQ Review Timeline

	MARYLAND	IOWA AND ALABAMA
Medical Records Requested	**7/91**	**10/91**
Record Review Feedback to Participants	**5/92**	**9/92**
Claims Profiles to Physicians	**5/92**	**9/92**
Physician Evaluation of Materials due back to DEMPAQ	**8/92**	**11/92**

What is participation?

Participation means that physicians will receive confidential feedback comparing their typical practice patterns to those of their peers. The basis for the feedback will be a medical record audit and a profiling of Medicare claims.

In addition, participating physicians will receive free category I CME credits if they complete a specified study of the record review results and respond with an evaluation of the review's accuracy, relevance and usefulness.

Finally, since physician feedback will be used to further refine the review methods, participating physicians have the opportunity to provide input into the development of methods and techniques which may be used in the future by HCFA to review physician office practice.

How does one participate?

Letters inviting participation will be mailed to a random sample of physicians who treated at least 25 Medicare patients during 1989 – 1990. Letters will first be mailed to a sample of Maryland physicians and then to samples of Iowa and Alabama physicians in the fall. While doctors are encouraged to participate, it is not mandatory.

Participating physicians will be asked to provide records for a randomly selected sample of approximately 25 patients to DEMPAQ for review. All information submitted is strictly confidential and will be reviewed by project staff only. The records will be photocopied, either by the physician or, if the physician so requests, by the PRO. To cover the costs of processing and mailing the records, a payment for each photocopied record will be offered.

How will feedback occur?

The exact feedback format is still being developed. We expect that it will include graphs and other pictorial comparisons of individual versus peer group performance. The detailed data used to create the graphs will be available upon request.◊

CLAIMS PROFILING

The claims profiling component of DEMPAQ complements the record review. Claims profiling offers the potential for a non-intrusive method of reviewing the quality of medical care. A medical record review requires the active participation of the physician in the review process. Claims profiling makes use of existing data sources to develop indicators of the quality of office-based care.

Claims profiling also provides a comprehensive view of medical care. By their nature, record reviews must be performed on a sample of charts. Statistical techniques then make it possible to generalize from the sample data to the entire population of a state. Claims data, on the other hand, cover all physicians who delivered care to all Medicare patients. Thus claims profiles give a view of all physicians and patients in a state.

Claims data may have limitations, however. Since they were developed primarily to pay physicians they may not contain information that is necessary to evaluate quality. One of the objectives of the DEMPAQ project is to compare the claims data to the medical record data as a source of information for measuring quality. Even if the claim can be used to appraise quality, the dimensions or aspects of quality it can evaluate may be quite different from those that can be evaluated through a record review. Thus claims data provide a broad view of medical care as compared to a record review which provides an in depth picture.

As with the record review the claims data are being summarized to provide confidential feedback to physicians about their practice patterns compared to those of their peers. The particular questions to be addressed by the claims profiling will be developed with input from members of the medical and data communities. Issues of interest to physicians in one state may be different from those in another. By soliciting input in the design of questions the nature of the profiles may be tailored to each participant state.◊

Record Review Criteria, con't from page 1

Drug Function

- Drug dose and frequency appear in medical record
- Drug is prescribed for appropriate condition
- No condition contraindication appears in medical record
- No drug contraindication appears in medical record
- Appropriate monitoring actions are taken

Follow-up

- If patient is hospitalized or referred then a discharge or consult note, or MD mention of event is in medical record
- Any mass, bleed, or unexplained weight loss was followed up in the time specified in the medical record, or within three months

Prevention

- Blood pressure exam is done once in two years
- Influenza inoculation is done once each winter season (October - February)
- Breast exam is done once for women in the two year review period
- Mammogram is done once for women less than 75 years old in the two year review period

Procedures

- Procedure is ordered/performed for appropriate indications
- There is appropriate follow-up of abnormal results
- Biopsy is done if required
- Pathology is in medical record if biopsy taken
- Area to be visualized by 'scope' is seen

Test Function

- Patient is hospitalized or referred if required
- Test is ordered/performed for appropriate conditions
- There is appropriate follow-up of abnormal results

Record Review Criteria, con't on page 4

Record Review Criteria, con't from page 3

Rates of Performance

To provide feedback to physicians about their *patterns* of practice means that DEMPAQ will calculate rates of performance. Rates describe overall performance instead of focusing on specific instances of care. A rate will be calculated for each Indicator within the seven Functions. For example, one of the proposed Indicators for Drugs is "Drugs prescribed for appropriate conditions." The denominator of this particular rate would be the number of times a course of drug therapy was prescribed; the numerator, the number of times where this prescription was appropriate to the patient's diagnosis or condition.◊

DEMPAQ Quarterly is published on a regular basis to promote greater understanding of the DEMPAQ project.

Comments or questions should be addressed to:

Ann G. Lawthers, Sc.D, Research Director
QCaRE, Harvard School of Public Health
677 Huntington Avenue
Boston, Massachusetts 02115

or

Jean E. Edwards, Project Manager,
Associate Executive Vice President
Delmarva Foundation for Medical Care, Inc.
9240 Centreville Road
Easton, Maryland 21601

DEMPAQ Quarterly
9240 Centreville Road
Easton, Maryland 21601

BULK RATE
U.S. Postage
PAID
Permit No. 82
Easton, MD

Inside this Issue of *DEMPAQ Quarterly*:

Record Review Criteria pg. 1
Record Review Timeline pg. 2
Claims Profiling pg. 3

DEMPAQ QUARTERLY

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 1, Number 4 September 1991

DEMPAQ is a three-year collaborative project to develop methods to review the care given to Medicare beneficiaries by physicians with office-based practices. The project focuses on education and feedback to physicians. Two review methods are being developed — medical record review and claims profiling.

RECORD REVIEW UPDATE

DEMPAQ Consultants and Criteria Development

Since March, DEMPAQ Physician Consultants have been reviewing draft criteria for the DEMPAQ project. Named by their specialty societies, the DEMPAQ Consultants are well respected leaders of the profession. While their participation in the project does not constitute endorsement by themselves or their organizations, they have provided thoughtful and useful input about the project's proposed criteria.

Consultants include members nominated by the American Medical Association, the American Academy of Family Physicians, the American Society of Internal Medicine, the American Osteopathic Association, the American College of Physicians, The Medical and Chirurgical Faculty of Maryland, the Maryland Academy of Family Physicians, the Maryland Society of Internal Medicine, the Iowa Medical Society, the Iowa Osteopathic Medical Association, the Iowa Clinical Society, the Iowa Academy of Family Physicians, the Medical Association of Alabama, the Alabama Society of Medicine, and the Family Practice Association of Alabama. The Consultants represent a cross-section of the physician groups most likely to be reviewed.

DEMPAQ Physician Consultants were asked to review draft criteria for the seven Functions covered by the Clinical Performance Assessment (*see box*) and criteria for the Record Documentation Assessment. The criteria were mailed in four installments; each installment covering one or more Functions. The criteria covered a range of diverse topics, from the appropriate indications for a test or a drug prescription, to appropriate workup items for a complaint.

DEMPAQ CLINICAL PERFORMANCE ASSESSMENT FUNCTIONS

Responding to a ***complaint***
Making a ***diagnosis***
Ordering a ***test***
Performing a ***procedure***
Prescribing a ***drug***
Follow-up of serious health threats
Providing ***prevention/screening*** services

Typically, quality assessment tools rely heavily on judgments made by individual peer reviewers. In such systems, the exact criteria being used are hidden from all except the reviewer (implicit criteria). In the DEMPAQ project, by developing detailed written criteria the judgments are made explicit so that physicians may determine for themselves their level of confidence in the criteria as measures of quality.

Record Review Update, con't on page 2

Linda E. Clark, RN, Executive Vice President
Delmarva Foundation for Medical Care, Inc.

R. Heather Palmer, M.B., B.Ch., S.M. (*British equivalent of MD*), Director
Center for Quality of Care Research and Education, HSPH

Record Review Update, con't from page 1

The process of developing criteria treads a fine line between over-simplification and excessive complexity. To provide feedback which promotes quality, the project must deal with complexity: medical practice is intrinsically difficult, and requires attention to many details. In the DEMPAQ project, we try to cope with complexity by focusing review upon the most common situations that could make a difference in patient outcomes.

DEMPAQ's interactive process opens the "black box" that traditionally surrounds criteria development. As stated in a recent *JAMA* editorial, "black boxes" used to review the quality of medical care should be opened and evaluated (Iezzoni, *JAMA* 1991; 265(22): 3006-3007). The DEMPAQ process provides the opportunity to evaluate the specific workings of the system both before and after it is tested. The DEMPAQ project is an evaluation of the record review technology.

DEMPAQ's process for developing the record review items opens new horizons for formulating criteria. The process is one of repeated interaction between the project research team and the Physician Consultants. The goal of such a process is to generate discussion about the validity and usefulness of individual criteria items. Field testing in each of the three states will lead to further refinement of the criteria.

At present, the criteria are being revised based on the first round of comments and suggestions from the Physician Consultants. Overall, the feedback from Consultants has been enormously helpful and has lead to substantive revisions in the criteria. Following a discussion of these revisions with the Consultants, the criteria will be tested in Maryland. After the Maryland test, we expect the need for further changes will emerge before the criteria are field tested in Iowa and Alabama next winter.

Record Review Computer System

The DEMPAQ Review System (DRS), a computerized data entry system, has been developed by project programmers using Paradox, a powerful

Record Review Update, con't on page 3

OTHER AMBULATORY REVIEW PROJECTS

HCFA has funded two other ambulatory review projects: one through the Wisconsin Peer Review Organization and one through the Massachusetts PRO (MassPRO). These projects and the DEMPAQ project do not compete with but complement one another. Brief descriptions of each project are presented here as information about other efforts to test and evaluate ambulatory review methods.

Wisconsin Ambulatory Review Project (WARP)

The Wisconsin Ambulatory Review Project is spearheaded by the Wisconsin PRO. It has designed a special data collection system to collect a variety of information about patient risk factors, therapeutic interventions and patient outcomes. This data system will be used to evaluate different therapeutic intervention strategies based on patient outcomes, adjusted for the patient's initial condition. Data on the effectiveness of different strategies will then be fed back to physicians. Peer review will be focused on physicians whose patients experience unexpectedly poor outcomes, taking into account patient risk factors.

Medicare Clinic Pilot Project

The Medicare Clinic Pilot Project is a joint venture between the MassPRO, the Cleveland Clinic, the Lahey Clinic and the Ochsner Clinic as well as the Ohio PRO (Peer Review Systems) and the Louisiana PRO (Louisiana Health Care Systems). The goal of the program is to assess the effects of treatment interventions and peer review on the outcomes of care. The project focuses on two diseases that are common among the elderly: diabetes and hypertension. The multispecialty clinic setting (Cleveland, Lahey and Ochsner Clinics) provides an integrated inpatient/outpatient service delivery setting for monitoring the interventions and patient outcomes. During the first phase of the project, the specific methods to be used in the implementation phase will be developed.

Record Review Update, con't from page 2

relational database. A reviewer takes data abstracted from the medical record and enters it into the computer. The computer applies the criteria to the data and generates profiles of physician practice.

DRS has several features which make it especially suited to the DEMPAQ project. In order to guide the reviewer through the medical record with the greatest efficiency, the computer screen presents a series of questions about the record being reviewed. These questions are "general" or "generic" questions which apply to all the *Functions* being reviewed. The generic questions form the framework for the review. The questions are accompanied by instructions which direct the reviewer in searching for the information necessary to answer the question.

Example of generic instructions: "Look ahead in the medical record and indicate if any of the following conditions are found."

The computer screen will then display a list of conditions following the above instructions that are specific to the item being reviewed. For example, to answer the question as to whether the medical record shows an appropriate reason for ordering a test, the reviewer must know which test was ordered. Acceptable reasons for ordering a hematocrit are, of course, different than those for ordering a serum glucose.

The text, which varies depending on the clinical item being reviewed, is contained in a *dictionary* file. Each Function has a separate dictionary, e.g., Test, Procedures, Complaints, etc. The specific text contained in the dictionaries can easily be modified to accommodate changes in the criteria, without having to change the entire computerized data entry system. This feature enhances the flexibility of the system. Criteria can readily be changed, added, or removed.

Thus DRS combines a generic framework with specific criteria. Using Paradox for the structure of the generic review system ensures that the system will be able to be modified in the future by other programmers. Paradox is a widely used database program with many users. A programmer familiar with Paradox will be able to make substantive changes in the structure of the system, should those become necessary at some point in the future. Changes can be made to both the generic structure of the review system (by programmers) and the specific criteria (by users of the system).

UPDATE: CLAIMS PROFILING

Draft formats for the two types of claims profiling, generic and focused, are in the process of being developed. Once draft formats have been prepared, and sample data run through the profiles to demonstrate how they work, the formats will be sent to Data Consultants for review and comment. Comments will also be sought from representatives of each of the three study states. Physicians in each state may have particular questions or issues they wish to see addressed by profiles. The review and comment process is designed to seek such input.

DEMPAQ Quarterly is published on a regular basis to promote greater understanding of the DEMPAQ project.

Comments or questions should be addressed to:

Ann G. Lawthers, Sc.D, Research Director
QCaRE, Harvard School of Public Health
677 Huntington Avenue
Boston, Massachusetts 02115

or

Jean E. Edwards, Project Manager,
Associate Executive Vice President
Delmarva Foundation for Medical Care, Inc.
9240 Centreville Road
Easton, Maryland 21601

DEMPAQ and Feedback

The cornerstone of the DEMPAQ educational philosophy is feedback to physicians. The DEMPAQ project collects information about physician office practice so group profiles of quality of performance may be fed back to physicians. This philosophy is different from the traditional PRO review process which has historically emphasized individual physician compliance with specific criteria.

While DEMPAQ uses criteria in the review process, it is not with the goal of identifying outliers. DEMPAQ's goal is to collect information about aggregate performance and aggregate patterns of care, as related to selected indicators of quality. This information is intended to spur efforts to improve quality.

DEMPAQ has enlisted the American Medical Research and Review Center (AMRRC) to work with the researchers in developing an educational package to accompany the profiles that will be fed back to physicians. The package will combine graphical and narrative materials to illustrate the results of the review. For this phase of the project, physician participants will be asked to review both the validity of the criteria and the usefulness of the feedback materials. In addition, physician participants will receive profile information comparing their performance to that of their peers.

DEMPAQ Quarterly
9240 Centreville Road
Easton, Maryland 21601

BULK RATE
U.S. Postage
PAID
Permit No. 82
Easton, MD

Inside this Issue of *DEMPAQ Quarterly*:

Record Review pg. 1
Other Ambulatory Review Projects pg. 2
Update: Claims Profiling pg. 3
Continuing Medical Education pg. 4

DEMPAQ QUARTERLY

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 2, Number 1 January/February 1992

DEMPAQ is a three-year collaborative project to develop methods to review the care given to Medicare beneficiaries by physicians with office-based practices. The project focuses on education and feedback to physicians. Two review methods are being developed — medical record review and claims profiling.

DEMPAQ IN THE FOREFRONT OF DEVELOPING METHODS FOR PROFILING PHYSICIANS' PRACTICE USING CLAIMS DATA

The claims profiling component of the DEMPAQ project is one of its key features. The goal of this component is to develop new and innovative ways to extract information from the Medicare claims database for evaluating the quality of ambulatory medical care. This task has brought many challenges. As detailed in this article, using Medicare claims data in analyses of quality has many advantages, and at the same time, several notable disadvantages.

The Medicare Claims Database: Potential and Pitfalls

The Medicare Claims database used to generate the profiles of physician performance has features that make it a potentially useful tool for monitoring quality of care. First, claims data are readily available. Unlike chart review data, which require special data collection efforts, Medicare claims data have already been collected and are available from the Health Care Financing Administration (HCFA). Because the data already exist, claims review is a non-intrusive way to evaluate quality.

Claims data also present a view of the entire population, not just a sample. While statistical techniques make it possible to generalize from a sample, claims provide a view not only of all Medicare beneficiaries but also of all the physicians who delivered their care. It is possible to link Part B claims data (physician visits, laboratory tests, and other ambulatory care information) with Part A data (institutional or hospital care).

This linkage affords a nearly complete picture of a patient's use of medical services. (Some services, such as drug prescriptions, are not covered by Medicare and are not included in the picture.) Moreover, the Medicare claims database is a national database, and all claims incurred by a patient are available, regardless of where the service was rendered.

Disadvantages to using claims data as a quality monitoring tool, relate to the fact that this database was designed for accounting rather than research purposes.

DEMPAQ in the Forefront, con't on page 2

ADVANTAGES OF USING CLAIMS DATA TO MONITOR QUALITY

- **Claims data readily available**
- **Non-intrusive review methods**
- **Population focus rather than sample focus**
- **Claims capture almost all service utilization**

Linda E. Clark, RN, Executive Vice President
Delmarva Foundation for Medical Care, Inc.

R. Heather Palmer, M.B., B.Ch., S.M. (*British equivalent of MD*), Director
Center for Quality of Care Research and Education, HSPH

DEMPAQ in the Forefront, con't from page 1

It was developed to allow payment of charges by the federal government, not to evaluate quality. A particular item's accuracy depends on its importance for billing. Items needed to pay a provider are generally very accurate; others, less so.

Medicare claims data present many special technical and administrative challenges. It is not an "easy" database to use. Claims data in the form available from HCFA are not ready to analyze. A single claim does not represent a single visit or a single laboratory test. Rather, one claim may be an aggregate of many "transactions." To be useful, claims data must be combined with elements from other databases. This process of combining data is complicated, e.g., a "clinically relevant episode" of care may involve multiple claims submitted by multiple providers across a several month time period. The claims database also contains many "extra" claims, such as duplicate records or records used for administrative purposes only.

Finally, and perhaps most relevant to quality monitoring efforts, claims lack the clinical detail of the medical record. This does not mean that claims cannot be used to monitor quality. Special efforts must be made to devise innovative methods to extract clinically meaningful information from the claims. The DEMPAQ project has made this task one of its foremost goals.

DEMPAQ Analysis Strategy

Analyzing the Medicare claims data involves several tasks. First, the project identifies primary care physicians from the database. These physicians have specialty codes representing general practice, family practice and internal medicine. Next, each patient (beneficiary on the claim) is assigned to the primary care physician who provided most of that patient's care (as indicated by claim activity) during the year being analyzed by the project (July 1, 1990, through June 30, 1991). A patient for whom a physician provides the majority of care over the course of a year is called a "primary patient."

To help account for differences in physicians' casemix, DEMPAQ will use Ambulatory Care Groups or ACGs. ACGs were developed at Johns Hopkins University and are intended to partially account for differences in health service use based on a person's age, gender, and pattern of disease over time (Weiner JP, et al., "Development and Application of a Population-Oriented Measure of Ambulatory Care Case-mix," *Med Care 1991*; 29(5): 452-472). All patients will be assigned to an ACG, and a weighted average of these ACGs will be calculated for each primary care physician's practice.

Once these initial steps have been accomplished, the data are structured into patient-level profiles which describe physician practice patterns and performance. DEMPAQ has developed formats for three types of profiles: office-practice profiles, condition-specific profiles, and preventive care profiles. The office-practice profiles summarize data about the physician's case load (primary vs non primary patients), distribution of ACGs, visits by patients to this and other doctors, distribution of visits (brief, extended, etc.), and primary patient characteristics (age and gender distribution).

Condition-specific profiles present data for a selected sample of patients with certain conditions. The project limits the profiling to patients with diabetes, hypertension, chronic obstructive pulmonary disease, rheumatoid or osteo arthritis, ischemic heart disease, or congestive heart failure. The condition-specific profiles present information about the use of services that contribute to good management of patients with each of the above mentioned conditions. The profiles also show use of inefficacious or contraindicated types of care. The indicators for these profiles are being developed in conjunction with input from clinicians.

The preventive care profile examines rates for certain preventive services, such as influenza immunizations, sigmoidoscopy, colonoscopy, and pap smears and mammograms for women.◊

CHALLENGES FACING THE USE OF CLAIMS DATA TO MONITOR QUALITY

- **Data originally developed to pay bills, not to monitor quality**
- **Accuracy of data items varies**
- **Technical and administrative challenges**
- **Lack of clinical detail**

RECORD REVIEW CRITERIA DEVELOPMENT: AN UPDATE

The last newsletter briefly described the criteria development process used by DEMPAQ. What follows describes in more detail the philosophy behind the development of the review criteria used in the computerized DEMPAQ Review System (DRS).

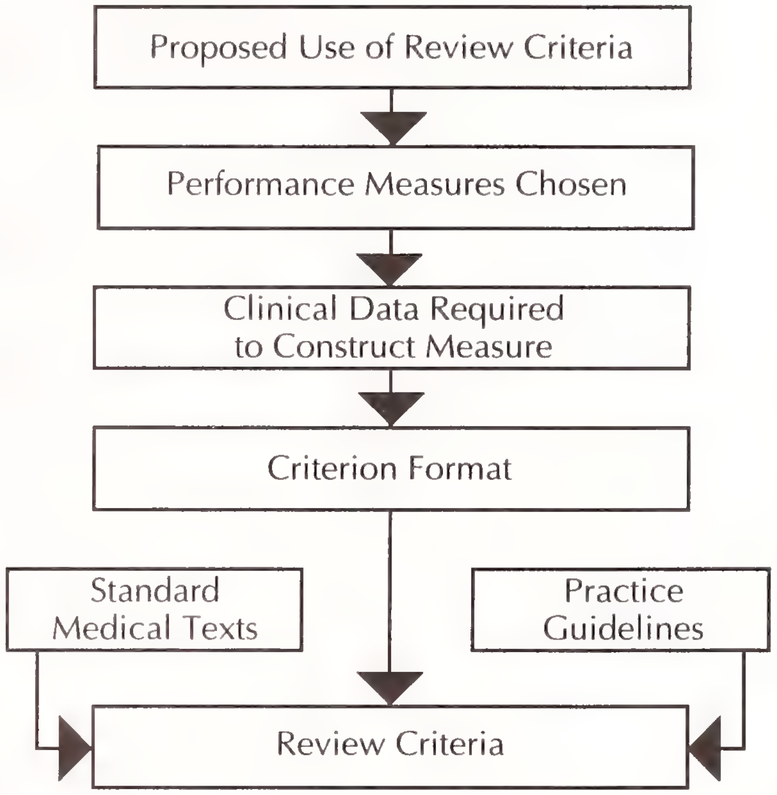
Before the review criteria can be developed, several hard questions need to be asked: How will the results of the review be used? Whose performance will be measured and what types of *performance measures* are required? The DEMPAQ project's philosophy is to provide feedback for education and provider-led quality improvement. Since an individual provider's performance must be measured over many functions, rates (indicators) of performance by function would be appropriate. Rates also have the advantage of being easily understood and explained, qualities that enhance their usefulness in education. Thus the proposed use of review criteria dictates the type of performance measures that need to be constructed.

The next step in developing criteria is deciding what types of *clinical data* are required to construct the performance measures. An early decision was that one of the major sources of data to measure performance would be the ambulatory care, or office, record. Since physicians in the ambulatory setting perform many tasks or functions, DEMPAQ decided to construct performance rates at the level of functions: e.g., diagnoses, complaints, drugs, tests, procedures, prevention and follow-up activities. DEMPAQ refers to categories of tasks as "Functions" (capitalized). Examples of the Functions are called clinical items. Clinical items supply the raw data necessary to develop the rates.

Once the types of clinical data needed have been identified, the specific formats for collecting them must be developed. *Criterion formats* are generic worksheets for collection of the clinical data used in scoring the performance measures. DEMPAQ used the generic worksheet approach in developing its criteria for each clinical item. First, categories of criteria were outlined for each Function. For example, drug indications became a generic category of criteria for the drug Function. Appropriate responses to a complaint became a category for the complaint Function.

After general criterion formats were developed, the assignment became one of entering specific *review criteria* on the worksheets. To accomplish this task practice guidelines were consulted, where available. Standard medical texts were also referenced. The draft review criteria were then circulated to DEMPAQ Physician Consultants for review, comment and revision.◊

DRS CRITERIA DEVELOPMENT PROCESS


Proposed Use of Review Criteria
Performance Measures Chosen
Clinical Data Required to Construct Measure
Criterion Format
Standard Medical Texts
Practice Guidelines
Review Criteria

DEMPAQ Quarterly is published on a regular basis to promote greater understanding of the DEMPAQ project.

Comments or questions should be addressed to:

Ann G. Lawthers, Sc.D, Research Director
QCaRE, Harvard School of Public Health
677 Huntington Avenue
Boston, Massachusetts 02115

or

Jean E. Edwards, Project Manager,
Associate Executive Vice President
Delmarva Foundation for Medical Care, Inc.
9240 Centreville Road
Easton, Maryland 21601

RECORD REVIEW: AN UPDATE

The record review will be conducted in all three study states: Alabama, Iowa and Maryland. Review is well underway in Maryland with over 2,000 ambulatory records received by the Delmarva Foundation for Medical Care. By mid December almost 700 of the 2,000 charts had been reviewed. Delmarva reviewers report that the computerized record review system (DRS) developed by the project is working very well.

Both the Iowa and Alabama PROs have mailed letters inviting physicians to participate in the DEMPAQ project in their states. Review of charts in those states is anticipated to begin in late winter.◊

PEER REVIEW

The peer review procedure for the DEMPAQ project actually includes two separate but related processes. First, all physicians participating in the project will peer review their own cases. To receive CME credit, they will be asked to complete forms summarizing the review of their "flagged" cases. The forms will direct the review with specific questions geared towards evaluating the review criteria.

The second part of the peer review process is a random review of 200 records by nonparticipating physicians. These peer reviewers conduct a directed re-review of selected cases. They will record their answers on forms similar to those used by participants. Their review will be conducted in the presence of a nurse reviewer familiar with the DRS system who may answer questions concerning the software and the review process.◊

DEMPAQ Quarterly
9240 Centreville Road
Easton, Maryland 21601

BULK RATE
U.S. Postage
PAID
Permit No. 82
Easton, MD

Inside this Issue of *DEMPAQ Quarterly*:

DEMPAQ in the Forefront pg. 1
Record Review Criteria Development pg. 3
Record Review: An Update pg. 4
Peer Review pg. 4

DEMPAQ QUARTERLY

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 2, Number 2 Spring 1992

DEMPAQ is a three-year collaborative project to develop methods to review the care given to Medicare beneficiaries by physicians with office-based practices. The project focuses on education and feedback to physicians. Two review methods are being developed — medical record review and claims profiling.

DEMPAQ PEER REVIEW: A STRUCTURED EVALUATION OF ITS RESEARCH TOOLS

This spring the DEMPAQ project began field-testing of its medical records review and will begin testing of the claims profiling procedures in the early summer. The field-testing consists of a detailed, structured peer review of both processes. Unlike peer review conducted by many state peer review organizations (PROs), which focus on detection of errors of individual physicians, DEMPAQ peer review is primarily a research tool designed for evaluation of the method of performance measurement. DEMPAQ peer reviewers are physician volunteers drawn from all primary-care specialty societies in each of the three study states: Maryland, Iowa, and Alabama.

Peer Review of the Medical Records Review Process

A unique feature of the DEMPAQ peer review process is its goal of evaluating the criteria used to review physicians' ambulatory care records. As part of the ongoing process to improve the record review, DEMPAQ uses peer review to evaluate the correctness of the criteria for record review or abstracting. In addition to assessing compliance with criteria for care, the peer review process also looks for possible quality-of-care problems that may be identified from the records.

Peer reviewers were recruited from the state branches of the American Society of Internal Medi-

A Structured Evaluation, con't on page 2

GLOSSARY

Claims or Practice Profile:
Summary description of a physician's practice pattern developed from Medicare claims for his/her patients.

Component or Criterion:
A predetermined element of care against which aspects of the quality of care provided may be compared.

Flag:
A computer-generated note made during initial DEMPAQ review (abstraction) to indicate possible noncompliance with a criterion of care.

Initial Reviewer or Abstractor:
Nurse trained by local PRO to review (abstract) physicians' medical records for individual patients using the DEMPAQ review system.

Peer Reviewer:
A physician member of a primary-care specialty society who evaluates the DEMPAQ review system by examining the records review and claims profiling processes.◊

Linda E. Clark, RN, Executive Vice President
Delmarva Foundation for Medical Care, Inc.

R. Heather Palmer, M.B., B.Ch., S.M. (*British equivalent of MD*), Director
Center for Quality of Care Research and Education, HSPH

A Structured Evaluation, con't from page 1

cine, the American Academy of Family Practice, and the American Medical Association. The peer reviewers are credentialled and trained by the PRO in their state. Their task is to examine a randomly chosen sample of records flagged by the initial DEMPAQ review (data abstraction) because a review criterion appeared to be unmet. For example, DEMPAQ criteria recommend that diabetic patients have an annual examination of the ocular fundi (funduscopy). The record of a diabetic patient would be flagged if the initial reviewer saw no indication that the fundi had been examined during the time covered by the review.

Figure 1 shows how the peer review process operates. To conduct the review, the peer reviewer uses a set of worksheets with the review questions and the criteria for the flagged item plus the medical record. First, the accuracy of abstraction of data from the record is assessed. The peer reviewer then is asked to look at the particular circumstances of the case:

- Do the data in the medical record justify the flag?
- Are there unique circumstances (e.g., patient factors, practice factors, or extenuating circumstances) which mean that the criterion could not be met in this case?
- Did the error result from unclear abstracting rules?

The answers to these questions define the limitations of the review criteria. In the last step of the evaluation process, the peer reviewer considers whether the review criterion for a particular clinical item (e.g., an indication for the diagnosis of diabetes) is correct. If the criterion is incorrect, the reviewer is asked to suggest a revised criterion.

The example of recommended annual funduscopy for diabetic patients can serve to illus-

A Structured Evaluation, con't on page 3

PEER REVIEW EVALUATION OF MEDICAL RECORDS

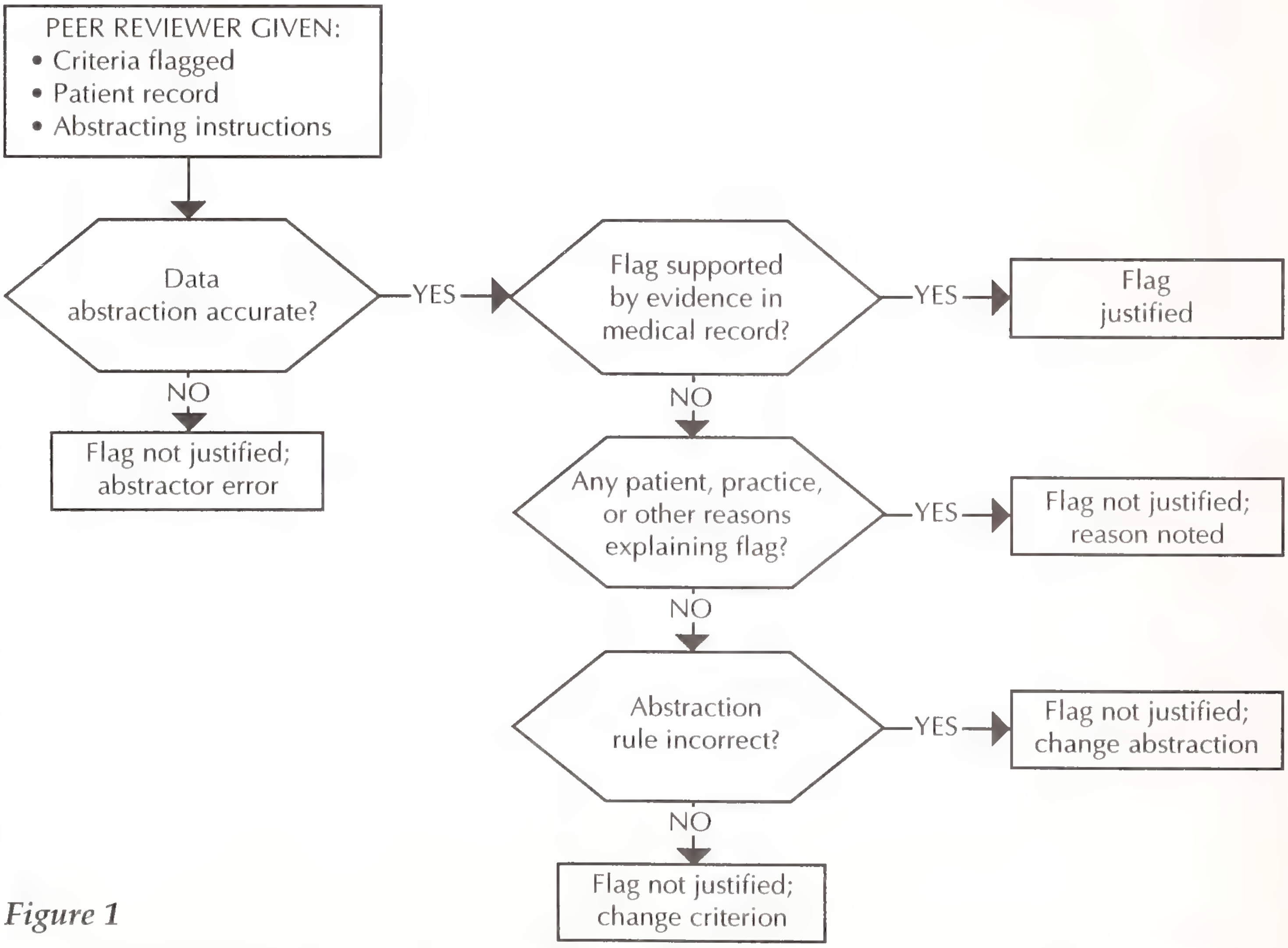

PEER REVIEWER GIVEN:
• Criteria flagged
• Patient record
• Abstracting instructions
Data abstraction accurate?
YES
NO
Flag not justified; abstractor error
Flag supported by evidence in medical record?
YES
Flag justified
NO
Any patient, practice, or other reasons explaining flag?
YES
Flag not justified; reason noted
NO
Abstraction rule incorrect?
YES
Flag not justified; change abstraction
NO
Flag not justified; change criterion

Figure 1

The Summer 1992 issue of the *DEMPAQ Quarterly* will feature articles on forthcoming activities:

- **Evaluation of claims profiling components**
- **Reporting to participating physicians on profiling of their claims data and review of their records**
- **Testing the accuracy of claims data for quality review**

A Structured Evaluation, con't from page 2

trate how the peer reviewer might assess a flag. If the peer reviewer learned from reading the patient's record that she had terminal cancer, failure to perform the annual examination of the ocular fundi could correctly be judged as "Criterion not met because of patient factors." The physician's judgment that the eye examination was unnecessary because of the patient's limited life expectancy supersedes the criterion. The peer reviewer would briefly explain these circumstances in the Comments section of the worksheet.

Finally, the peer reviewer is asked to comment on the adequacy of documentation for the case flagged. Unless the physician has noted the reason why a recommended care item was not performed, it maybe impossible to tell whether the omission was justified.

Peer reviewers in Maryland began their work in mid-April after orientation by staff of the Delmarva Foundation for Medical Care. In Iowa and Alabama, peer review will begin in the early summer. Since DEMPAQ investigators anticipate that the peer reviewer may have questions about both the initial review process and the DEMPAQ criteria, a nurse-reviewer who is familiar with both the review process and the records flagged will be present at all peer review sessions to answer questions about review of the individual records as well as the criteria.◊

RECORDS REVIEW UPDATE: PHYSICIAN RECRUITMENT AND REVIEW STATUS

Review of office practice records of participating Maryland physicians is now complete. Ninety-one physicians took part in this aspect of the study, and 1750 records were reviewed. The Alabama Quality Assurance Foundation, Inc., the collaborating PRO in that state, reports that 100 physicians in Alabama have agreed to participate and have already submitted 1900 records for review. At the beginning of April, 200 of these records had been reviewed. Recruitment of physician participants is continuing in Iowa, where by early April 55 physicians had submitted 1033 records for review by the Iowa Foundation for Medical Care.

Reports from the local PROs indicate that the review process is going smoothly in all three states. Reviewers have found that the DEMPAQ review system (DRS) is easy to use, and they have been able to complete their reviews efficiently. One of the Delmarva reviewers described her experience working with DRS as follows: "The DRS computer system has helped simplify a complex, tedious review. Easy access to screens allows me to enter pertinent information quickly and accurately and to retrieve the clinical data selections in order to reaffirm the choices when necessary. Having had no computer experience, I found the program to be very user-friendly."◊

***DEMPAQ Quarterly* is published on a regular basis to promote greater understanding of the DEMPAQ project.**

Comments or questions should be addressed to:

Ann G. Lawthers, Sc.D, Research Director
QCaRE, Harvard School of Public Health
677 Huntington Avenue
Boston, Massachusetts 02115

or

Jean E. Edwards, Project Director,
Vice President for External Relations
Delmarva Foundation for Medical Care, Inc.
9240 Centreville Road
Easton, Maryland 21601

CLAIMS PROFILING UPDATE

All data needed for the claims profiling have been received. Much thought has been given to determining the best format for presentation of claims profiles. Proposed draft formats have been sent to several consultants, and their critiques and suggestions are being incorporated into the final format. The claims profiling reports will be sent to two groups: the collaborating state medical societies and individual physician consultants to the DEMPAQ project.

Each physician consultant will receive eight separate profile reports detailing specifics of his/her practice. These will include one report each for office-practice patterns and preventive care and separate reports for care of each of six specific conditions (congestive heart failure, chronic obstructive pulmonary disease, diabetes mellitus, hypertension, ischemic heart disease, and osteoarthritis). Profile reports will consist of descriptive text as well as tabular presentation of data. A feature of the profiling is that each consultant physician will receive an overall rating of performance of individual care items (e.g., the hemoglobin A1C test for diabetic patients) and a comparison of his/her performance to that of all physicians in the state.

The group profile reports sent to participating state medical societies will include information on all physicians in the state; data on physician performance will be listed by specialty and possibly by region. Like the individual reports sent to consultants, the group reports will include profiles of office practice patterns, preventive care, and condition-specific care. Use and distribution of these reports will be at the discretion of the society.

DEMPAQ staff expect that profile reports along with reports on the record review will be mailed to participating physicians in the fall of 1992.

DEMPAQ Quarterly
9240 Centreville Road
Easton, Maryland 21601

BULK RATE
U.S. Postage
PAID
Permit No. 82
Easton, MD

Inside this Issue of *DEMPAQ Quarterly*:

DEMPAQ Peer Review: A Structured Evaluation of Its ResearchTools pg. 1
Records Review: Physician Recruitment and Review Status pg. 3
Claims Profiling Update pg. 4

DEMPAQ QUARTERLY

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 2, Number 3 Summer 1992

DEMPAQ is a three-year collaborative project to develop methods for review of care given to Medicare beneficiaries by physicians with office-based practices. The project focuses on education and feedback to physicians. Two review methods are being developed — medical record review and claims profiling.

DEMPAQ PEER REVIEW: A STRUCTURED EVALUATION OF ITS RESEARCH TOOLS, Part 2

Peer review, as implemented by the DEMPAQ project, is used to evaluate methods developed by the project to assess physicians' performance. The spring issue featured a report on peer review of the DEMPAQ review system (DRS), a process for computerized review of medical records that was recently completed in Maryland. Researchers have now turned their attention to peer review of claims profiling components, which will be conducted by the Maryland PRO in the fall. This activity is described in the following article.

Peer Review of Claims Profiling Components

Evaluation of the claims profiles by peer reviewers will complement the DEMPAQ peer review of the medical records review process. The peer review evaluation of claims profiling components is designed to answer the question:

Is each clinical item or component in the profile a valid measure of the quality of medical care?

DEMPAQ claims profiles selected for inclusion in the peer review are the prevention-services profile and three condition-specific profiles: diabetes, chronic obstructive pulmonary disease (COPD), and congestive heart failure (CHF). Profile components fall into one of two care classifications: recommended and limited use. Recommended care represents the minimal level of care that should be provided to a patient with a given condition. Limited use care includes items whose use depends on the patient's condition and the physician's discretion.

A Structured Evaluation, con't on page 3

GLOSSARY

Claims or Practice Profiling:
Profiling is an analytic tool that uses epidemiologic methods to compare practice patterns of providers on the dimensions of cost, service use, or quality (process and outcome) of care.

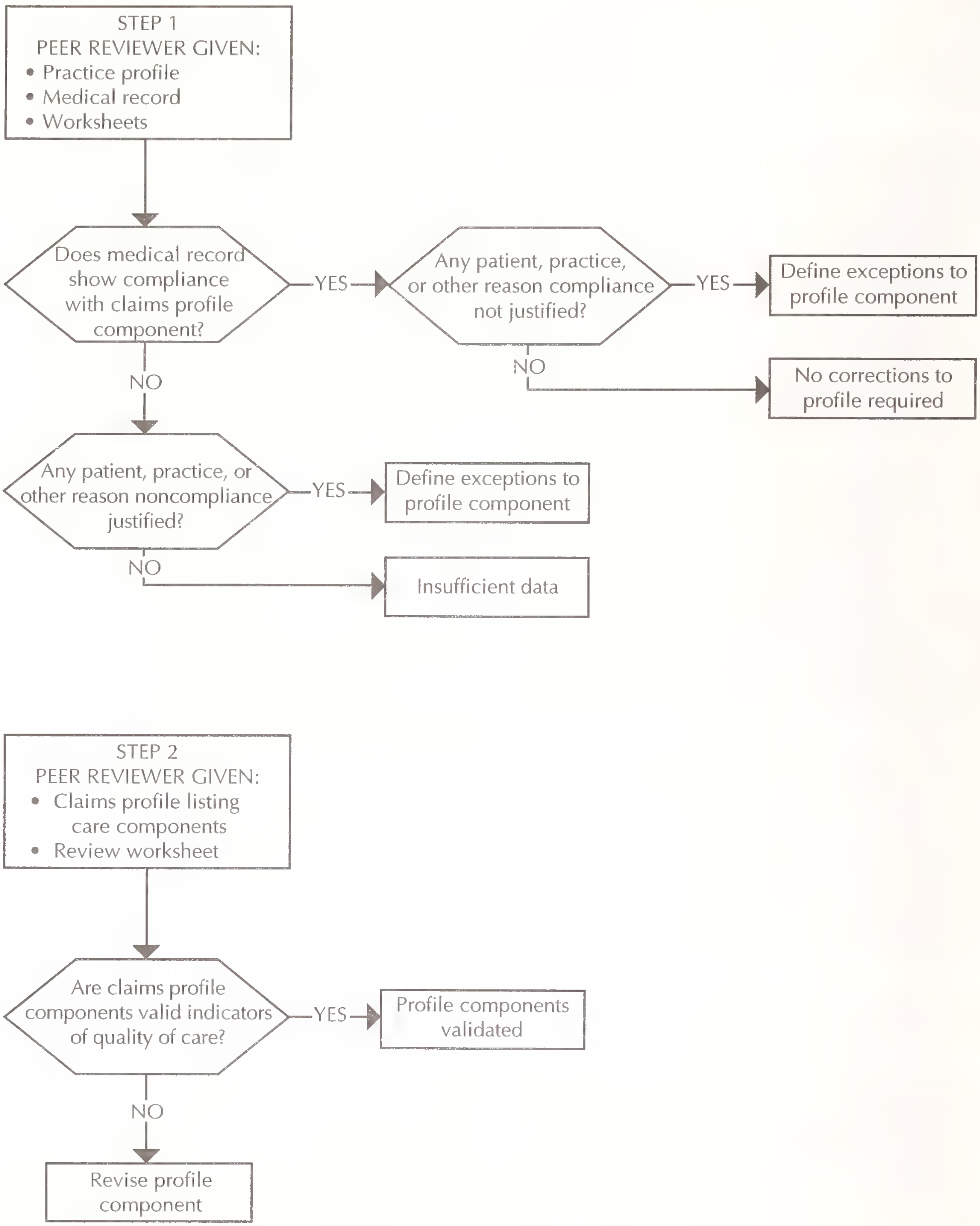
Component or Criterion:
The term *component* or *criterion*, as used by DEMPAQ researchers, signifies a predetermined element of care against which aspects of the quality of care provided may be compared.

Flag:
A flag is a computer-generated note made during the initial DEMPAQ review (abstraction) of the medical record that indicates possible noncompliance with a criterion of care.

Peer Reviewer:
A peer reviewer is a physician member of a primary-care specialty society who evaluates the DEMPAQ review system by examining the records review and claims profiling processes.◊

Linda E. Clark, RN, Executive Vice President
Delmarva Foundation for Medical Care, Inc.

R. Heather Palmer, M.B., B.Ch., S.M. *(British equivalent of MD)*, Director
Center for Quality of Care Research and Education, HSPH



PEER REVIEW EVALUATION OF CLAIMS PROFILES

STEP 1
PEER REVIEWER GIVEN:
• Practice profile
• Medical record
• Worksheets
Does medical record show compliance with claims profile component?
YES
Any patient, practice, or other reason compliance not justified?
YES
Define exceptions to profile component
NO
No corrections to profile required
NO
Any patient, practice, or other reason noncompliance justified?
YES
Define exceptions to profile component
NO
Insufficient data
STEP 2
PEER REVIEWER GIVEN:
• Claims profile listing care components
• Review worksheet
Are claims profile components valid indicators of quality of care?
YES
Profile components validated
NO
Revise profile component

Figure 1

A Structured Evaluation, con't from page 1

Figure 1 (page 2) shows the sequence of queries that guide the peer reviewer through the evaluation of the claims profiling components. The reviewer uses three tools to complete the profile component evaluation: the claims profile, the medical record, and DEMPAQ peer review worksheets. Peer reviewers will go through a two-step process to conduct their evaluations. The first step deals with issues specific to a particular case. For example, in the case of a recommended-care item that was performed, the peer reviewer will make judgments on whether giving that care was appropriate for the patient and whether any factors were present to suggest that the care should not have been done. These factors, as noted in the medical record, will be described by the peer reviewer. If recommended care was not given, the peer reviewer will search the medical record to determine whether an adequate reason for the omission was documented and will describe the reason on the worksheet. The second step of peer review will address the validity of the measures and the appropriateness of the classification (e.g., recommended care or limited-use care).

The peer reviewer's task can be described with the diabetes profile as an illustration. One component of this profile is an annual visit to an ophthalmologist. If, in considering the specifics of a particular case, the peer reviewer discovers from the medical record that the patient (whose claims data contain no mention of an ophthalmology visit) had refused the recommended consultation, he/she may conclude that the component was not met due to patient factors. After reviewing the records of all diabetic patients, the peer reviewer addresses the issue of face validity. For example, he/she is asked to respond to the question of whether the recommendation that diabetic patients see an ophthalmologist once each year is appropriately included in a profile of physician performance that uses Medicare claims data.

During the peer review session, any questions that the peer reviewer may have on the medical record, profile, or worksheets can be addressed to the nurse reviewer. Data on completed worksheets will be analyzed by DEMPAQ researchers.◊

RELIABILITY STUDY MEASURES AGREEMENT BETWEEN CLAIMS AND RECORD REVIEW DATA

In the DEMPAQ Reliability Study, researchers are determining the usefulness of claims data as a measure of the quality of ambulatory care. An important part of this study is a comparison of information in the medical record with information in a Medicare claim. The Reliability Study will try to answer several questions:

- To what extent does the information in the claim agree with the information in the record?
- Does the extent of agreement vary by patient characteristics (e.g., diagnosis or age)?
- Does the extent of agreement vary by site of care (e.g., physician's office versus emergency room?)

To address the issue of agreement, the researchers developed a specialized computer program, which was used by reviewers at the Delmarva Foundation for Medical Care to abstract key data elements (e.g., visits, diagnoses, and results of laboratory tests) from the medical record. Abstraction of several thousand records has produced a data file based on medical records that the researchers are now comparing with the data file developed from Medicare claims. Among the analyses planned is an examination of data at the event-specific level. This analysis will look at issues such as whether date and diagnosis reported in the claim and record agree. Knowing the degree of accuracy of the claims data is essential to this type of research because the claims profiles would be undermined if the database from which they are constructed contains inaccuracies. Results of this study are expected to be available by the end of the year.◊

DEMPAQ Quarterly **is published on a regular basis to promote greater understanding of the DEMPAQ project.**

Comments or questions should be addressed to:

Ann G. Lawthers, Sc.D, Research Director
QCaRE, Harvard School of Public Health
677 Huntington Avenue
Boston, Massachusetts 02115

or

Jean E. Edwards, Project Director,
Vice President for External Relations
Delmarva Foundation for Medical Care, Inc.
9240 Centreville Road
Easton, Maryland 21601

RECORDS REVIEW UPDATE: DRS PEER REVIEW COMPLETED IN MARYLAND

Peer review of the Maryland medical records review process was completed in mid-July. The seven physicians who served as peer reviewers held 18 sessions with the nurse reviewer at the PRO office. During a three-month period, this group completed review of a sample of over 300 items that had been flagged during the initial review process, which included approximately 1,700 records submitted by 91 physicians. The DEMPAQ review system (DRS) was lauded by the peer reviewers as being an impressive tool with the potential of providing valuable information on health care given to Medicare recipients in the ambulatory-care setting.

In Iowa the 53 participating physicians had submitted well over 1,000 records by mid-July, and most of these had been reviewed. Peer review was expected to begin there late in July, and DRS review was scheduled for completion by early August. In Alabama, where nearly 900 records submitted by 100 physicians had been reviewed as of mid-July, peer review is slated to begin in early fall. The Alabama PRO expects that DRS review of the submitted medical records will be completed by that time.◊

CLAIMS PROFILING UPDATE: SEVEN PROFILES COMPLETED

DEMPAQ researchers have finalized the contents of the six condition-specific profiles and the preventive-services profile. The six conditions being studied are congestive heart failure, ischemic heart disease, hypertension, chronic obstructive pulmonary disease, diabetes, and osteoarthritis. Each profile lists care items that should (recommended care) or may (limited use care) have been used by a patient with a given condition. The profiles also display resource use (e.g., hospitalizations and office visits) as well as timely follow-up in the ambulatory-care setting of patients who have been hospitalized. To explore variations in resource use among physicians, researchers will develop separate profiles for patient groups with different severity of illness.

DEMPAQ Quarterly
9240 Centreville Road
Easton, Maryland 21601

BULK RATE
U.S. Postage
PAID
Permit No. 82
Easton, MD

Inside this Issue of *DEMPAQ Quarterly*:

DEMPAQ Peer Review: A Structured Evaluation of Its ResearchTools, part 2 pg. 1
Reliability Study Measures Agreement between Claims and Record Review Data pg. 3
Records Review Update: DRS Peer Review Completed in Maryland pg. 4
Claims Profiling Update: Seven Profiles Completed pg. 4

DEMPAQ QUARTERLY

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 3, Number 1 Fall-Winter 1992–93

DEMPAQ is a three-year collaborative project to develop methods to review the care given to Medicare beneficiaries by physicians with office-based practices. The project focuses on education and feedback to physicians. Two review methods are being developed — medical record review and claims profiling.

CLAIMS PROFILING: METHODS AND OBJECTIVES

Physicians submit claims for reimbursement of costs incurred in the care of individual patients. Claims data can be analyzed to form profiles of physician performance (rates of performance). Physician profiles are a nonintru-sive way of evaluating the quality of medical care provided to their patients. The objective of the DEMPAQ project is the development and evaluation of profiles derived from both Medicare claims data and from review of physicians' office records of ambulatory care for the elderly. DEMPAQ's profiles based on physicians' practice records offer a detailed view of the care given to a small sample of Medicare beneficiaries, while profiles derived from claims data provide a complementary, population-based measure of quality of care. The databases used for DEMPAQ claims profiling include Medicare Part A, Part B, and Part B institutional data.

Profile Specifications

To measure quality of the care reported in Medicare claims, DEMPAQ researchers, in collaboration with the project's consultants, formulated six focused, condition-specific profiles and a prevention profile. These profiles specify a narrow set of care items that a patient with a certain condition should usually receive (recommended care) or should receive under certain circumstances (limited-use care), along with the clinical rationale for each care item. Next, the condition-specific profiles show the frequency of hospitalization for selected diagnoses that may be markers of poor ambulatory care. The profiles also allow researchers to track use of health-care resources and promptness of follow-up after hospital discharge.

The conditions selected for profiling represent diagnoses that cause significant morbidity in the elderly and thus include a sufficient number of cases in the data base to make claims profiling feasible. The six conditions profiled were chronic obstructive pulmonary disease (COPD), congestive heart failure (CHF), diabetes mellitus, hypertension, ischemic heart disease (IHD), and osteoarthritis. Table 1 (page 2) lists five care items recommended for diabetics: primary care visit, hemoglobin A1C test, urinalysis, ophthalmology visit, and influenza immunization. Additional recommended care items (not shown) are three blood lipid tests (total serum cholesterol, HDL-cholesterol, and triglycerides).

The office practice profile describes the use of ambulatory, hospital, and ancillary services associated with the practices of individual physicians and, at the statewide level, by all Medicare patients in the state. At the level of the individual practitioner, the practice profiles will report facts such as the total number of Medicare patients seen by the physician and, of the total, the number who received most of their primary care from that physician (primary patients) and their age distribution. Visits made by the physician's primary patients to other health-care providers (including specialists) will also be displayed, as will the number of various types of visits

Claims Profiling, con't on page 2

Linda E. Clark, RN, Executive Vice President
Delmarva Foundation for Medical Care, Inc.

R. Heather Palmer, M.B., B.Ch., S.M. (*British equivalent of MD*), Director
Center for Quality of Care Research and Education, HSPH

TABLE 1:
SOME RECOMMENDED-CARE ITEMS FROM THE DIABETES PROFILE

RECOMMENDED CARE	COMMENTS: Recommended care represents the minimal level of services that should be provided to a patient.
Primary care office (home) visit	According to the American Board of Family Practice, visits for diabetic patients in control should be scheduled every 3 – 6 months. This report shows the percentage of patients with diabetes who had a primary care visit at least once every 6 months.
Hemoglobin A1C test	Since the hemoglobin A1C test gives a good indication of the level of glucose control, the American Board of Family Practice suggests that it be done every 2 – 3 months. However, this profile recommends that the test be done at least once each year.
Urinalysis	Annual urinalysis is recommended to check for proteinuria. Proteinuria is an early manifestation of diabetic nephropathy.
Ophthalmology visit	A complete eye and visual examination should be performed annually. This examination is important to detect early signs of retinopathy or cataract.
Influenza immunization	Influenza immunization is recommended by the Immunization Practices Advisory Committee from the Centers for Disease Control and is particularly important for patients with chronic metabolic disease.

Claims Profiling, con't from page 1

(e.g., brief, intermediate, extended). Other health-care events such as hospital admissions and use of laboratory tests or x-rays will also be reported. At the statewide level, practice profile reports give cumulative data for all practitioners in the state who care for Medicare patients.

The final drafts of the profiles emerged after completion of a three-step process. First, a preliminary version of each profile was drafted based on suggestions of physicians in each of the three participating states. Second, these drafts were sent to nationally recognized data consultants and to a small number of physicians representing state medical societies for review and comment. Last, these comments were reviewed by the team of investigators and incorporated into the final drafts of the profiles along with citations to the literature that supported the recommendations.

Profile Reports

For the condition-specific and prevention profiles, the profile report lists data on the percentages of patients with each condition who received profile-specified care. Table 2 (page 3) shows the profile format for reporting the distribution of recommended services for patients with diabetes, while Table 3 (page 3) shows the format of the office practice profile giving the distribution of patients' ambulatory care visits to their primary care physician.

Evaluation of Profiles: Reliability and Validity

DEMPAQ investigators are testing the value of profiling as a tool for monitoring quality of care (see *DEMPAQ Quarterly* 2[3], 1992). The Reliability Study, designed to test the correspondence between claims data and information in the medical record, compared data from all of the roughly 2000 individual patient records submitted by physicians in Maryland with data in Medicare claims for the patients whose records were submitted. Among the findings that emerged were the following:

- For 27 diagnoses, agreement between claims and record ranged from 72% to 99%.
- For performance of 31 tests and procedures, agreement ranged from 70% to 99%.

Claims Profiling, con't on page 3

TABLE 2:
PROTOTYPE FORMAT FOR PROFILE REPORT SHOWING DISTRIBUTION OF SERVICES PROVIDED TO PATIENTS WITH DIABETES*

RECOMMENDED CARE	PERCENT OF PATIENTS WITH SERVICES			
	0	1–2	3–5	>5
General medical visits				
First six months	6.65	45.65	34.19	13.45
Second six months	15.50	37.39	33.05	13.98
Hemoglobin A1C test	76.29	16.56	6.35	1.10
Urinalysis	54.08	30.79	10.80	4.33
Optometry/Ophthalmology visit	51.84	31.65	11.95	4.55
Influenza immunization	89.18	10.81	0.01	0

**Data can be listed by specialty of practitioners or for all practitioners statewide.*

TABLE 3:
PROTOTYPE FORMAT FOR PROFILE REPORT TO AN INDIVIDUAL PHYSICIAN SHOWING DISTRIBUTION OF AMBULATORY CARE VISITS MADE BY HIS/HER MEDICARE PATIENTS

VISITS BY PRIMARY PATIENTS	AVERAGE NUMBER OF VISITS	PATIENT DISTRIBUTION BY VISITS				
		0 Visits	1 Visit	2–4 Visits	5–9 Visits	10+ Visits
Total	6.62	0	36	59	31	19
To you	3.54	0	24	37	19	4
To all others	2.08	0	12	22	11	15

Claims Profiling, con't from page 2

- Patient names and gender matched in ≥98% of cases, but date of birth was missing from 30% of the records.

In the Profile Component Evaluation Study, physician peer reviewers looked at records data to see if there were clinical reasons why recommended or limited-use care was or was not given. The study was limited to four profiles (CHF, COPD, diabetes, and prevention); for each profile records for 50 patients with the indicated condition (or, in the case of prevention, 50 Medicare patients) were compared with claims data. When the claims indicated that, for example, a patient with congestive heart failure had not received an influenza immunization, the physician consultant searched the record for evidence that might explain the absence. Valid reasons might be patient refusal, evidence in the record that influenza vaccine was given elsewhere, or terminal illness.

PHYSICIAN APPRAISAL OF DEMPAQ TOOLS

In October, R. Heather Palmer, M.B., B.Ch., S.M., DEMPAQ Principle Investigator, Jonathan Weiner, Sc.D., a Co-investigator, and Jean Edwards, Sandy Moentmann, and other staff from the Delmarva Foundation for Medical Care, met with Maryland physicians who served as consultants in the development of the project's quality-of-care criteria and those who conducted peer review both of DEMPAQ's medical review process and of the claims profiling components. Dr. Palmer presented some preliminary findings that emerged from the project's record review. Results presented at the meeting generated enthusiasm on the part of both researchers and consultants. Results of claims profiling were presented at a November meeting with the consultants. Again, the physician consultants expressed considerable interest in the preliminary findings on claims profiling.

The investigators are now preparing to provide results from both record review and claims profiling to the physician consultants in Iowa and Alabama. Later, results will be discussed with participating physicians in all three states.

DEMPAQ Quarterly is published on a regular basis to promote greater understanding of the DEMPAQ project.

Comments or questions should be addressed to:

Ann G. Lawthers, Sc.D,
Research Director
QCaRE, Harvard School of Public Health
677 Huntington Avenue
Boston, Massachusetts 02115

Jean E. Edwards, Project Manager,
Vice President of External Relations
Delmarva Foundation for Medical Care, Inc.
9240 Centreville Road
Easton, Maryland 21601

DEMPAQ Quarterly
9240 Centreville Road
Easton, Maryland 21601

BULK RATE
U.S. Postage
PAID
Permit No. 82
Easton, MD

Inside this Issue of *DEMPAQ Quarterly*:

Claims Profiling Update pg. 1
Physician Appraisal of DEMPAQ Tools pg. 4

DEMPAQ QUARTERLY

DEVELOPING AND EVALUATING METHODS TO PROMOTE AMBULATORY CARE QUALITY

Volume 3, Number 2 Spring 1993

DEMPAQ is a three-year collaborative project to develop methods to review the care given to Medicare beneficiaries by physicians with office-based practices. The project focuses on education and feedback to physicians. Two review methods are being developed — medical record review and claims profiling.

INVESTIGATORS REPORT TO PHYSICIAN CONSULTANTS ON FINDINGS FROM RECORD REVIEW AND CLAIMS PROFILING

Over the last several months, meetings were held with the physician consultants in each of the three DEMPAQ study states. Each had a common set of objectives (figure 1), which was reached through much useful and constructive discussion. The following briefly summarizes the content of each meeting.

Maryland Meeting

DEMPAQ investigators met with Maryland physician consultants in Baltimore on October 21, 1992. Participating physicians represented the Medical and Chirurgical Faculty of Maryland, the Maryland Society of Internal Medicine, and the Maryland Academy of Family Physicians. The meetings were hosted by the Delmarva Foundation for Medical Care, Inc. (DFMC), the collaborating PRO in Maryland.

Results of record review were reviewed in light of data from peer review and studies of inter-rater reliability (i.e., degree of agreement between two different medical abstractors). During the initial discussion, aimed to familiarize all participants with the record review criteria and data from the review, liaison physicians asked questions and expressed some of their concerns. For example, abstractors had been trained to interpret the timing requirement of *every 6 months* for tests or follow-up quite strictly as six months (plus/minus one month); it was suggested that the criteria be revised to *twice within the year, at least 4 months apart.*

As the meeting proceeded, physicians had the opportunity to identify interesting findings; they often recommended additional review of specific cases to clarify potential quality problems. The meeting served as a forum for open discussion about the results and their interpretation and demonstrated how the review results could be used to select topics for continuing medical education (CME) workshops.

The investigators met again with Maryland physician consultants on November 4 to discuss the data on claims profiling. The project's philosophy on the use of profiles to improve the performance of all physicians (figure 2, page 2) was presented and contrasted with more traditional uses of profiling. The claims profiles were then reviewed, along with the

Investigators Report, con't on page 2

OBJECTIVES OF DEMPAQ PHYSICIAN MEETINGS

- ❖ **Promote understanding of DEMPAQ methods**
- ❖ **Present record review and claims profiling results**
- ❖ **Communicate strengths and limitations of results**
- ❖ **Identify potentially interesting findings for further review or possible CME**

figure 1

DEMPAQ'S APPROACH TO PROMOTING QUALITY IN AMBULATORY CARE

- **Focus on quality**
- **Target groups, not individuals**
- **Obtain input from clinicians and methodologists**
- **Display patterns of practice**
- **Use explicit, publicly-disclosed criteria**
- **Provide educational feedback to groups**

figure 2

Investigators Report, con't from page 1

strengths and limitations of the data. Physicians expressed particular interest in the claims data as a means for summarizing all the care that a patient has received from many different sources. Encouraging findings were reviewed concerning the accuracy of diagnosis information in claims data. Physicians noted that, at least during the study period, it was not possible to review preventive care in the claims data because the Health Care Financing Administration (HCFA) did not then pay for preventive care. The general office practice profiles were found to give a graphic portrayal of the different styles of practice.

The liaison physicians said that they had acquired useful information about the practice of medicine from their experience as peer reviewers for DEMPAQ.

Alabama Meeting

On January 28 and 29, 1993, DEMPAQ investigators met in Birmingham with the project's physician consultants who represented collaborating medical groups in the state of Alabama. These included the Medical Association of Alabama, the Alabama Society of Medicine, and the Family Practice Association of Alabama. The meetings were hosted by the Alabama Quality Assurance Foundation, Inc. (AQAF), the collaborating PRO in Alabama.

The claims profiling component of the study was discussed at the initial session on January 28. After an introduction, the investigators led a discussion of the profiles developed by DEMPAQ for six chronic conditions that affect the elderly (see *DEMPAQ Quarterly*, Vol. 3, No. 1). Meeting participants offered their views on the reasonableness of the results and areas where further refinement of the tool will be necessary. Next, preliminary results of two substudies were presented. The claims reliability study compared data from individual patient records with Medicare claims data for those patients (see *DEMPAQ Quarterly*, Vol. 3, No. 1). The second substudy, an evaluation of profile components, sought reasons documented in the medical record that would explain the absence from claims data of a specific item of recommended care.

Before the meeting adjourned, individual practice profiles were distributed to all physician consultants present. As in Maryland, participating physicians in Alabama will receive feedback on the records review and claims profiling later this year.

On January 29, the investigators met with physician consultants and PRO representatives involved with record review. Discussion centered on DEMPAQ criteria for drug usage, diagnoses, tests, and management of the six chronic conditions profiled in the claims study. A number of ideas for revision of DEMPAQ criteria emerged. For example, in a discussion of criteria for use of antibiotics, it was stated that DEMPAQ criteria were perhaps too rigid as appropriate standards for prescribing these drugs, whose usage requires knowledge of community-acquired infections in the local area. Participants' suggestions for revision have been documented for future use in revision of the review criteria.

Iowa Meeting

On February 25 – 26, 1993, the investigators met with Iowa physician consultants in Des Moines. The Iowa Foundation for Medical Care, Inc. (IFMC), the collaborating PRO in Iowa, hosted the meetings. The format, agenda, and objectives were similar to the previous sessions. The sessions opened with an overview of the DEMPAQ methods and philosophy. Each meeting then devoted a substantial block of time to an explanation of the DEMPAQ tools and feedback report formats. Examples of each type of report (claims and records) were explained in detail, with time for questions.

Next, the discussion turned to the results of the review. As the results were presented, the physicians were asked to judge whether the result seemed believable (based on the results of DEMPAQ reliability and validity studies) and whether the finding was potentially interesting from a Continuing Medical Education (CME) standpoint.

As in the other two states, the discussion with the Iowa physicians was thoughtful and covered a wide range of topics. Many useful suggestions for improving the claims and records review methods emerged from the discussion. For example, it was suggested that influenza immunization should be eliminated from all profiles since HCFA rarely pays for flu shots. Regarding records, one of the suggestions involved removing "monitor diet" from the list of required monitoring actions for hypertension. Most of the participants felt that this item was of less clinical significance than other monitoring actions.

Investigators Report, con't on page 3

DEMPAQ RESEARCHERS PRESENT FINDINGS TO VARIED AUDIENCES

On October 15, 1992, Jonathan P. Weiner, Dr. P.H., DEMPAQ investigator and faculty member at Johns Hopkins University School of Public Health, gave a talk entitled "The Application of Claims Data to Quality of Care Profiling" at the Annual Medicare Carrier/Intermediary Medical Review Conference in Chicago. The meeting was sponsored by the National Blue Cross/Blue Shield Association.

Deborah W. Garnick, Sc.D., a member of the team at Harvard School of Public Health and Associate Research Professor at Brandeis University, discussed the development of the claims profiling methodology at the Chicago Blue Cross and Blue Shield Association's 1992 Medical and Quality Management Conference on October 21, 1992. Dr. Garnick's presentation, "Development and Application of Quality of Care Indicators," offered a preview of the initial results of the project's study of claims profiling.

Jinnet Fowles, Ph.D., a DEMPAQ investigator at the Park Nicollet Medical Foundation in Minneapolis, presented "Challenges in Practice-Based Research" at the Third Primary Care Research Conference convened by the Agency for Health Care Policy and Research in Atlanta, January 10-12, 1993. Dr. Fowles focused on the results of the DEMPAQ claims reliability study (see "Investigators Report to Physician Consultants on Findings from Record Review and Claims Profiling," page 1). She pointed out that data of this type are needed because research on primary care is based on a variety of information sources whose accessibility, reliability, and validity are unknown.

On November 5, 1992, Ann G. Lawthers, Sc.D., DEMPAQ investigator at Harvard School of Public Health, presented the claims profiling aspect of DEMPAQ before a national audience. The talk, entitled "Using Medicare Data to Develop Profiles of Primary Care Physicians: The DEMPAQ Project," was given at the annual meeting of the American Public Health Association, held in Washington, D.C.

On January 29, 1993, Drs. Palmer and Lawthers made a presentation entitled "Developing and Evaluating Methods to Promote Ambulatory Care Quality" at the Lister Hill Center for Health Policy, University of Alabama at Birmingham. Both the claims profiling and record review components of the project were presented, along with some preliminary data for the states of Maryland and Alabama.

Investigators Report, con't from page 2

Next Steps

The discussions in each state helped identify areas where the review instruments and review processes could be improved. These recommended changes will be transmitted to HCFA as part of a report on the feasibility of the DEMPAQ approach to review of general medical practice in Medicare. The results of the meetings in all three states will also be summarized. These reports will underscore the conviction that it is possible to work cooperatively with physicians for quality improvement.

Physician participants who submitted up to 25 Medicare records to the DEMPAQ project will receive the results of the medical records review and claims profiling activity by mail in the summer and fall of 1993. They will also be invited to attend a CME workshop. This workshop will focus on teaching physicians how to review their own records using the DEMPAQ workbook; how to improve quality of care; and what the scientific basis of the DEMPAQ criteria is (see "DEMPAQ Contract Modified to Support CME Workshops," page 4).

Acknowledgments
Contributors in the state of Maryland: Christian E. Jensen, M.D., M.P.H., DFMC; Sheldon Goldgeier, M.D.; Christine Marino, M.D.; Abdul Nayeem, M.D.; Alfred Ossman, M.D.; Howard Wilson, M.D.; and Joseph Zebley III, M.D.; Jean Edwards, DFMC; Sandra Moentmann, DFMC; Judy Brower, DFMC; and Nancy Eber, DFMC.

Contributors in the state of Alabama: Robert Sherrill, M.D., AQAF; Laurie Hall, M.D.; George C. Smith, Jr., M.D.; Pink Folmar, M.D.; Brian Perry, M.D. and Peter Morris, M.D.; Steve Boger, M.D.; Nadine Schiesz, AQAF; Carlene Gibbons, AQAF; and Barbara Baites, AQAF.

Contributors in the state of Iowa: James Maroc, M.D., IFMC; John Carroll, M.D.; Kevin Quinn, M.D.; Milt VanGundy, M.D.; Larry Severeidt, M.D.; Dana Shaffer, D.O.; Kevin deRegnier, D.O.; Robert Good, D.O.; Jay Rosenburger, D.O.; Steven Craig, M.D.; Norm Pawleski, IOMA; Donna Bottroff, IMS; Jenny VanGundy, IFMC; Nancy Paine, IFMC; Becky Hemann, IFMC; and Sherri Segebart, IFMC.◊

***DEMPAQ Quarterly* is published on a regular basis to promote greater understanding of the DEMPAQ project.**

Comments or questions should be addressed to:

Ann G. Lawthers, Sc.D, Research Director
QCRE, Harvard School of Public Health
677 Huntington Avenue
Boston, Massachusetts 02115

or

Jean E. Edwards, Project Director,
Vice President of External Relations
Delmarva Foundation for Medical Care, Inc.
9240 Centreville Road
Easton, Maryland 21601

DEMPAQ CONTRACT MODIFIED TO SUPPORT CME WORKSHOPS

HCFA has recently made a contract modification to cover the costs of a DEMPAQ Continuing Medical Education (CME) workshop in each study state. The workshops, to be held in the summer and fall of 1993, will disseminate results of the study to the physicians who submitted records for review, as well as to other interested physicians. Contacts have been made with individuals at the University of Maryland and the University of Alabama about plans for workshops in these states. Soon contacts will be established in Iowa.

The proposed CME effort is part of the overall DEMPAQ objective of testing strategies for quality improvement. The CME workshops would communicate the importance of quality monitoring and management in daily clinical practice, teach participants how to use the DEMPAQ tools for self-review and education, and present detailed information about patient management and practice organization for selected topics.

The conversations with each of the physician groups began to highlight potentially interesting findings that might serve as potential CME topics. Based on our meetings in all three states, the following lists possible topics for CME:

- Hypertension management and the use of newer antihypertensive drugs
- Diabetes management and controversies about monitoring glucose control and monitoring for complications
- Follow-up of abnormal test results
- COPD diagnosis and use of bronchodilators
- Ischemic heart disease diagnosis and admission for arrhythmias
- Congestive heart failure
- Prevention, especially breast exam and mammogram
- Antibiotic use

In addition to selecting potential CME areas, we also plan to work on ways of presenting the findings using case reviews that will make it easier for physicians to understand.◊

DEMPAQ Quarterly
9240 Centreville Road
Easton, Maryland 21601

BULK RATE
U.S. Postage
PAID
Permit No. 82
Easton, MD

Inside this Issue of *DEMPAQ Quarterly*:

Investigators Report pg. 1
DEMPAQ Researchers Present Findings pg. 3
DEMPAQ Contract Modified pg. 4

Chapter 4
Focus on Quality: Profiling Physicians' Practice Patterns

Abstract

Federal and private policy makers now recognize the potential of profiling physicians' practices as a component of continuous quality improvement. To help meet the need for claims-based profiling methods that can be applied to improve the quality of care delivered to Medicare patients, we formed a multidisciplinary team of physicians, health services researchers, a biostatistician, computer programmers and staff of three Peer Review Organizations (PROs). Our goal was to develop new methods that PROs could use to profile the ambulatory care provided by primary care physicians to Medicare patients in three states.

This chapter is divided into two main sections. In the first section, we define profiling more fully, point out some advantages and disadvantages associated with claims-based profiling to improve the quality of physicians' practice and propose a set of principles to guide the development of quality focused profiles. In the second section, we describe the system we developed based on those principles to profile the care primary physicians provide to Medicare patients.

Introduction

Federal and private policy makers now recognize the potential of profiling physicians' practices as a component of continuous quality improvement (Shapiro et al., 1993, PPRC 1992, Lasker 1992, Findlay 1993). Specifically, profiles of physicians' practices developed with a focus on quality could be used to:

- Inform physicians about how closely the care they provide conforms to their peers' practice patterns;
- Track adherence to practice guidelines; and
- Measure the quality of care provided by groups of physicians, including members of physician networks, staff of HMOs, or providers who serve Medicare patients in a community.

However, profiling methods already developed to meet cost containment goals will not be useful for quality improvement goals.

To meet cost containment goals during the 1980s, health insurance companies developed methods to "profile" physicians' practices. Using claims data, the insurers identified providers with expensive practice patterns so that they could be sanctioned. For example, physicians who consistently used more resources than their peers would suffer financial penalties or would be cut from an insurer's network. In some managed care plans, efficient providers were rewarded. For instance, physicians whose patterns fell within established norms would "bypass" the requirements for detailed case-by-case review or would be selected to participate in preferred provider networks (Stephenson 1992, Nathanson et al., 1993). These uses of profiling as a cost containment tool have become widespread (Findlay 1993).

By 1990, insurers representing 96% of the Preferred Provider Organization (PPO) market reported that they maintained physician profiles (Hoy et al., 1993). In 1992, 45.5% of physicians were subject to either clinical or economic profiling, although less than half were regularly provided with feedback (Emmons and Wozniak 1993).

In the current environment, new profiling methods are needed to improve the quality of physicians' services. A simple example shows why physicians' practice profiles to be used for quality improvement must differ from profiles emphasizing cost. The Medicare program now compiles comparative performance reports which are used to alert physicians whose claims submission rates for selected services are higher than the norm for their specialty (Robinson 1990). One measure is the number of colonoscopy claims a physician submits per 100 beneficiaries he or she treats.

For quality improvement purposes, this rate alone is insufficient because it is subject to multiple sources of bias:

- The denominator may be inflated by patients who are incidental to a physician's practice.
- The feedback of a raw rate is not currently linked to a physician's practice.
- The rate may be artificially low for primary care physicians who refer almost all their patients to gastroenterologists for colonoscopies.
- The raw rate is not clinically meaningful; that is, it is not adjusted to account for the specific case-mix characteristics of patients in each physician's practice.
- The emphasis on individual physicians stresses the assigning of blame rather than the improvement of the performance of all physicians.
- There are no substantial educational strategies for increasing clinical knowledge or improving performance.

Our goal was to develop new methods that PROs could use to profile the ambulatory care provided by primary care physicians to Medicare patients in three states (the team also developed a new method to review office-based medical records using explicit criteria {Lawthers et al., 1993}). This effort is among the first, and perhaps the widest in scope, ever to develop claims-based profiles focused on quality rather than the cost of care (another profile development effort focused on quality is described in Leatherman et al., 1991).

This chapter is divided into two main sections. In the first section, we define profiling more fully, point out some advantages and disadvantages associated with claims-based profiling to improve the quality of physicians' practice, and propose a set of principles to guide the development of quality focused profiles. In the second section, we describe the system we developed based on those principles to profile the care primary physicians provide to Medicare patients.

Background

Definition of Physicians' Practice Profiling

Profiling is an epidemiological technique that focuses on patterns of care rather than on individual occurrences of care, usually using large computerized databases (PPRC, 1992).

There are several disadvantages to profiling as a quality improvement tool. Large data sets developed to pay bills often lack clinical specificity, test results, and information on medication—limitations that can be overcome by linking profiling with record reviews of a sample of cases (Weiner et al., 1990).

Despite these shortcomings, claims profiles offer many advantages over other review methodologies. Five specific advantages are:

- While claims data often lack the desired clinical depth of information, they can provide a broad view of activity. Unlike records-based tools, these profiles can readily incorporate information about all services that a beneficiary receives, linking information about episodes of care across time, providers, and settings.
- Because claims have already been collected as part of the billing process, they are much less expensive to use, given the volume of information available.
- Claims-based profiles do not require intrusion into physicians' practices. They represent an electronic audit trail of the interaction between the health care provider and consumer. Profiles represent a nonintrusive way to assess providers' performance, thus reducing the

"hassle" factor. A far larger number of physicians' practices can be screened using this broad based approach than by using more expensive reviews of medical records for selected individuals.

- Claims databases can be analyzed for varying time intervals and with varying aggregations of providers. If small samples for a particular condition are a problem, it is possible to increase the number of observations by increasing the length of the study period.
- Claims profiles are flexible; as criteria for care evolve, components of the claims based profiles can be updated.

Principles for Developing Physician Profiles

To be a useful tool for improving quality of care, physician practice profiles need to:

- Be flexible so profiles can be easily changed as medical practice changes;
- Involve the end user (practicing physicians) in developing profiles;
- Include an explicit assessment of the reliability of the profile; and
- Be fair to various groups of providers.

Flexibility. Profiles of physicians' practices must be designed to easily incorporate changes in treatment for specific conditions based on new research findings, newly developed technologies, and newly published practice guidelines. Currently, if a practice guideline calls for a HgbA1c test for all patients who have diabetes, then the profile can be designed to present data on the percentage of patients with the condition who receive the test. In the future, if a better test is developed, a new procedure code (CPT-4) is registered, and the Health Care Financing Administration (HCFA) pays for the test, the profile can be revised to reflect these changes.

If a profile using 1991 data shows that in one area of a state the rate of use of the HgbA1C test is very low, then educational efforts about the guideline would most efficiently be targeted to physicians in that area. Rather than show the data aggregated to the level of all physicians in the state, the profile could be revised to show data for specific areas where the claims data or other quality measurement tools show that there might be the need for intervention.

Finally, profiles that are designed to incorporate flexibility could be used to investigate in detail how guidelines are implemented. For example, profiles could be tailored to examine whether most physicians raise their level of testing to 80% of patients with the condition, or whether some physicians remain at a very low level of testing, while others test virtually all of their eligible patients. Moreover, if a revised guideline shows a test to be unnecessary, then profiles can be used to monitor physicians' incorporation of this new information into their practices.

Profiles also must be designed with the ability to incorporate improvements in data systems and coding. For example, more specificity of procedure coding was implemented in the beginning of 1992 along with the phase-in of physician payment based on resource based relative value scales (RBRVS).

Another example of the need for flexibility in profiles arose during the development and testing phase of our project. When we discovered very low rates of blood glucose and cholesterol tests, we found that it was caused, in part, by our initial omission of panels which are groups of 6, 12, 18, or 20+ tests, often grouped by condition, e.g., heart panel (personal communication, Judy Barr, Northeastern University).

An example of improvement in data systems is the implementation of unique provider identification numbers (UPINs) by HCFA. The profiles of physicians' practices we report in this paper are based on physicians' billing numbers, not UPINs. Under the previous system, physicians might have used several numbers if they practiced in different locations. Furthermore, several physicians in a group might have billed under the same number. Under the new UPIN system, all of a physician's services are billed under one number which represents only one physician. This modification will allow us to include in revised profiles, specific physicians who were only identified as members of a group under the old system.

User involvement. Development of physicians' practice profiles should be an ongoing process that involves users of profiles in all stages—initial design of profiles, interpretation of test data, improvement of earlier versions, interpretation of final versions, dissemination of profiles to practicing physicians, and designing educational interventions based on profile results. In the past, however, profiling systems often have fallen far short of this ideal because including users increased the time and expense of developing profiles.

Physicians are the key users of the information from profiles developed for quality improvement goals. They use the information to improve their own practices and the level of care in their communities. Physicians who feel ownership of the process of developing and disseminating information will be more apt to feel that the information is important. Moreover, they know best what information will be clinically meaningful and can pinpoint important issues for their patients. In contrast, insurers are the main users of the information from profiles developed for cost containment.

An explicit assessment of reliability. The usefulness of practice profiles depends, in part, on understanding how reliable the information is. While it is not necessary that the data be completely accurate, it is necessary to understand the direction and extent of the errors and whether the errors are systematic or random. The concept of reliability rests upon the reproducibility of the measure, for example, whether items are measured the same way at different points in time or by different people. For Medicare claims, one component of reliability is the accuracy of coding, which can be tested by examining the correspondence of claims information with the primary physician's ambulatory medical record (for further details, see Fowles et al., 1993).

Because claims are a "by-product" of the system of care rather than a deliberate clinical notation, physicians are suspicious about their accuracy. A careful assessment of the degree of data accuracy will address these concerns and allow physicians to focus on the utility of the information.

Fairness. To be fair to providers, physicians' profiles should include only those patients for whom that provider bears the major responsibility. These profiles should compare similar groups of providers (those with the same specialty, for example) and compare similar groups of patients (by selecting patients with the same condition and controlling for case-mix). However, meeting these conditions of fairness is often difficult.

Thorny issues can quickly arise even with a test for which there is a clear and well-accepted standard of care, such as the use of annual pap smears for certain categories of women. For example, unless one is profiling physicians in a health system where patients are assigned to a gatekeeper physician, including all patients for each physician's profile can introduce a serious bias. Consider two primary care physicians who each care for 1,000 women and perform pap smears on 700 of them: a rate of 70%. However, one physician also covers for several colleagues' vacations and sees an additional 150 women over the course of the year for one or two visits. If all 1,150 women seen are used as the denominator, this physician's rate of pap smears is only 61% (700/1,150). Thus, a profile should exclude those patients for whom a different physician provides a great deal of care.

In the case of pap smears, it would seem fair to compare rates across internists and family practitioners because the ideal rate approaches 100% (if the hysterectomy rate varies by region, then the optimal pap smear rate should also vary by region) without regard to provider specialty. However, if the internists in a community serve women with a socioeconomic, educational, or ethnic status that would make them more likely to keep appointments, while family practitioners serve women who may be less likely to comply with the physicians' recommendations, then any interpretations of a higher performance rate for internists than for family practitioners should take into account these differences in patient populations.

When the standard of care is less clear, the issue of comparing practice patterns across physicians is even more complicated. Consider, for example, the number of visits per year for a patient with chronic hypertension. The physician's practice style should be considered along with clinical factors, socioeconomic factors, benefit plan design and financial access which all can influence the pattern of care. Some physicians may routinely see their patients once each quarter while others may see their patients less often if their blood pressure is well controlled by medication. For this example, there may be a range of patterns of care, all of which are acceptable.

Developing Methods for Profiling Physician Practices with a Focus on Quality

In this section, we describe the steps we took to develop new approaches to physician practice profiling which included:

- Obtaining and preparing data;
- Determining which patients to include;
- Specifying the content of reports;
- Producing profiles; and
- Disseminating and evaluating profiles.

Obtaining and Preparing Data

Health insurance claims data are not ready to analyze immediately when the tape or cartridge of data arrives from Medicare. Rather, in order to prepare for data analysis several key steps are required:

- A useful unit of analysis (such as a visit to a doctor) should be developed from a set of several discrete financial transactions;
- Data elements from other databases (such as beneficiary files or provider files) should be combined with claims records; and
- Many superfluous claims (e.g. duplicate transactions) should be identified, removed, or modified (for further details, see Chapter 5).

We obtained the National Claims History (NCH) file from HCFA. While the NCH suggests it is only one file by its title, it is actually a series of three files linked by beneficiary ID. These files and a provider file are described briefly below.

Beneficiary file (HISKEW). This file contains socio-demographic information regarding the beneficiary including:

- Birth date;
- Sex;
- Type of Medicare coverage (e.g. over age 65 vs. disability or renal disease);
- Place of residence (zip code and county); and
- Death date (if applicable).

Office-based claims file. This file contains claims for services performed in an ambulatory facility and paid through Part B of Medicare (although not all Part B claims are found in this file). An ambulatory facility can be a physician's or group practice's office or a patient's home. Critical variables in this file include:

- Date of service;
- Procedure code;
- Diagnosis code;
- Location of service;
- Type of service;
- Cost of service; and
- Provider code.

Institutional claims file. The file contains both Part A and Part B claims paid to institutions including hospitals (both inpatient and outpatient services), nursing homes, ambulatory surgery centers and mental health facilities. The same key variables in the office-based claims file are also available in the institutional records, with the addition of the diagnosis related group (DRG) and patients' hospital discharge status.

Provider file. In addition, we obtained a provider file from the carrier containing information describing physicians' characteristics such as specialty, board certification status and participation in a group practice.

From these data files, we selected the variables and the observations (patients, claims and providers) needed to develop profiles. For example, from the 447,145 patients on the raw beneficiary file in Maryland, we selected the 389,765 (87.2%) of patients who met the following criteria. These patients:

- Were age 65 or older, alive the whole year, and entitled to Medicare from July 1, 1990 to June 30, 1991;
- Lived in the state the entire year;
- Were not enrolled in an HMO;
- Received some Part B services; and
- Did not have nursing home claims for two consecutive quarters.

Using the Part B claims data, we eliminated data on 19.1% of records (1,631,466 out of 8,541,698 records) that showed:

- Duplicates;
- Services that could not be linked to a beneficiary;
- Nonexistent diagnosis and procedure codes; and
- Services that could not be linked to a physician.

For the remaining beneficiaries, we constructed person level analytic files that contained all of their claims for inpatient or outpatient services for the entire year.

Determining Which Patients to Include

The goal of the primary care practice profile is to include those patients for whom the physician is responsible and exclude patients who see the physician only incidentally. This relationship also is referred to as the patient's usual source of care (Franks et al., 1993). Consider the following patient scenarios:

- Patient A gets all care from Dr. Smith, an internist;
- Patient B visits Dr. Smith once but receives most care from a specialist;
- Patient C visits Dr. Smith once but receives most care from a different internist; or
- Patient D visits only a cardiologist.

To deal with this range of scenarios, we developed the concept of primary care source (PCS) for each patient (in a managed care setting in which each patient is assigned to a "gatekeeper" physician, this step is not required). Each patient was assigned to the primary care physician (internist, family practitioner, general practitioner) who provided more care than any other primary care physician. Using only claims for "face-to-face" visits, the primary care physician associated with the majority of services to a single patient was designated as that patient's primary care source (PCS). In the case of a tie, total charges were used to assign the PCS. Sixty-eight percent of the 389,765 beneficiaries in the state of Maryland were assigned to a PCS who provided approximately 87.6% of their primary care visits. We did not include beneficiaries who could not be assigned to a PCS (i.e., they never visited an internist, family practitioner, or general practitioner) or if they were assigned to a group practice.

Returning to the patient scenarios:

- Patient A (who gets all care from Dr. Smith, an internist) is assigned to Dr. Smith.
- Patient B (who visits Dr. Smith only once but receives most care from a specialist) also is assigned to Dr. Smith since a specialist cannot be a PCS.
- Patient C (who also visits Dr. Smith once but receives most care from a different internist) is assigned to the other internist.

- Patient D (who visits only a cardiologist) is not included in our profiling system since a primary care physician is not involved in his or her care.

Specifying Content of Profiles

Development of profile content was an iterative process:

- First, the research team developed initial draft profiles for six conditions common in the Medicare population, for office based practice in general and for preventive care.
- Next, clinical experts, the states' Peer Review Organizations, local physicians in Maryland (and the other two states), and experts in data analysis reviewed the draft profiles for clinical logic, usefulness to providers, and clarity.
- Finally, the profiles were revised based on these comments.

The profiles were designed to be flexible so that they can be further revised in response to users' comments, new areas of interest, or changes in data systems or coding.

We developed explicit criteria forms to facilitate discussion of which utilization or outcome measures to include in our profiles of patients with specific conditions. These forms list:

- The measure (e.g., HDL cholesterol test);
- The data source (e.g., the specific procedure codes—CPT 83718);
- Notes explaining the clinical rationale for inclusion, references to medical texts, recommendations of professional medical associations, and practice guidelines; and
- *Caveats* about the interpretation of the measure.

The criteria are presented in Chapters 6–8.

Producing Profiles

We developed three types of physician practice profiles: condition-specific profiles, office practice profiles, and preventive profiles. Each measure in the profile is expressed as a rate where the numerators are the number of patients who received services or experienced outcomes, aggregated over time (usually a year), and the denominators are the relevant patients from among those who were assigned to the PCS.

Condition-specific profiles. These profiles were developed for six conditions that were selected on the basis of their frequency in the Medicare population, the likelihood that poor ambulatory management would result in adverse outcomes, and the availability of relevant information in the claims data. These conditions included diabetes, hypertension, congestive heart failure, chronic obstructive pulmonary disease, ischemic heart disease, and osteoarthritis.

For each condition, we developed a series of measures (because Medicare does not cover prescription drugs, we could not develop performance measures involving drugs):

- *Recommended care items* are tests or actions that should be done for all patients with a given condition, such as an annual ophthalmologic exam for diabetic patients.
- *Limited use items* are tests or actions that are appropriate for only a subset of the patients with the condition. For example, not all congestive heart failure patients require a chest x-ray each year, while, for the subset of patients with comorbid chronic obstructive pulmonary disease, a chest x-ray is quite likely to be needed. Thus some of the performance measures need to be stratified by conditions that reflect the severity or comorbidity of the patient's illness.
- *Hospitalization measures* identify the frequency of hospitalization for selected complications of a condition (an adverse outcome). The conditions selected were those which could have been a consequence of inadequate ambulatory care.
- *Follow-up visit measures* track how soon a patient is seen in the ambulatory setting after hospital discharge.

In addition, we examined how often specialist physicians treated patients with each condition.

For example, the diabetes profile for the state of Maryland for family practice (Chapter 6) shows the condition-specific profile for patients whose PCS' are family physicians and who had at least one claim in the year for a physician visit for diabetes. The interpretation of the profiles is fairly straightforward. For instance, Table 1 shows that 83.3% of patients with diabetes did not receive a HgbA1c test from July 1, 1990, to June 30, 1991, although the American Board of Family Practice (ABFP 1990) recommends the test every two to three months and the American Diabetic Association (ADA, 1989) recommends the test every six months. Another 13.2% had the test once or twice in the year, 3.2% three to five times, and .03% over five times. During a feedback session, physicians in Maryland appeared to conclude that conformance to this ADA guideline showed room for improvement.

Office practice profiles. The elements of these profiles, which are shown in Chapter 8, can be summarized as follows:

- Characteristics of the physician's case load (e.g., the breakdown of primary vs. nonprimary patients);
- Primary patient characteristics (age and gender distribution);
- Case mix distribution;
- Visits by patients to this and other physicians;
- Distribution of visits (brief, extended, etc.);
- Patterns of ancillary test use (by this and other providers); and
- Hospitalization and emergency department use of the physician's primary patients.

To help account for differences in physicians' case mix, we used Ambulatory Care Groups (ACGs) developed at Johns Hopkins University to partially account for differences in health services use based on a person's age, gender and patterns of morbidity over time (Weiner et al., 1991). Based on the diagnoses occurring on the claims file over a one-year period, all patients are assigned to an ACG that reflects the seriousness and number of unique conditions that affect the patient. The primary physician's use of resources for his/her patient panel is adjusted by comparing actual use of resources to the resources that would be expected for patients with similar burdens of disease.

In addition, standard resource units are applied to services to enable overall comparisons of the intensity of practice style rather than using billed charges or paid charges. For example, we might wish to examine the use of various types of laboratory tests, each of which is priced differently depending on which laboratory performed the test, the region of the country, and the type of laboratory test. For inpatient information we based our standardized reimbursement on the rate paid per Diagnosis Related Group (DRG), and for outpatient procedures we used the HCFA resource based relative value scale (RBRVS) weights, ignoring geographic factor price differentials, multiplied by the approximate 1992 factor of $31 (when no RBRVS weight was available, we used the average reimbursement in Maryland for the procedure; for further details see Parente et al., 1993).

Prevention profiles. These profiles were developed to demonstrate the capability of the system to measure and report preventive care patterns. Measures include influenza immunizations, sigmoidoscopy, and colonoscopy, as well as pap smears and mammography for women. We were aware that, since Medicare did not pay for many of these items during the study period, these services would not be found on claims. Therefore, these profiles do not currently represent the actual care given to patients. We were aware, however, that Medicare was considering payment for these items in the future. Table 1 of this chapter shows a sample preventive care table.

Disseminating and Evaluating Profiles

We are using a mixed strategy to feed back preliminary results to physicians. First, physicians who were part of the development process were shown the profiles in a series of briefings in each state. Next, a special Continuing Medical Education (CME) workshop was conducted in each state during which participating physicians received profiles both for their specialty group and, for the office based profiles, for their own patient population (for the condition-specific profiles, many physicians do not have a large enough number of patients to present stable results; this is a concern even when all of a physician's patients with a condition are included in this analysis).

The accuracy of the claims data was evaluated by comparing the information on diagnoses and procedures on claims data to information contained in the medical record. This analysis, which is reported in full in Chapter 9, shows claims to be specific but not sensitive. That is, if a diagnosis appeared on a claim, the patient usually had the disease according to the medical record (for further details, see Chapter 10). However, patients also had diseases not recorded in the claims. In contrast, claims more often indicated tests and procedures than had been documented in the primary care physician's record (which is partially explained by the facts that we did not also have access to specialty physicians' records and that lab test claims did not include the billing number of the ordering physician).

For our objective, developing measures of quality of primary care by determining whether patients with a specific disease received services recommended for that disease, we concluded that the claims data are appropriate. While the claims data underestimate the prevalence of a disease, they provide a fair picture of conformance to treatment guidelines. If there is a bias, it is in the direction of indicating more often that physicians are conforming to recommended care because there probably are more cases missing from the denominator (number of patients with the condition) than from the numerator (number of services provided) (see Chapter 9).

In addition, we are examining the face validity of the profiles, whether the measures appear to make sense (Abramson 1990), by having physicians compare the interpretation of the profiling results against an interpretation they derive from the office record. For example, if the claim indicated that a patient with diabetes did or did not receive a HgbA1c test, then the physician reviewed the record to see if that care was appropriate. This review suggested that the presence of a claim for recommended care items in the condition specific profiles indicated good quality care. If the claim is not found, three categories of reasons appear to explain its absence:

- The patient didn't need the service;
- The service may have been provided outside the Medicare system, e.g. in a Veterans Administration hospital; and
- The patient may not have received a needed service.

Discussion

In this chapter, we described the four attributes of physicians' practice profiles that make the profiles useful for quality improvement—flexibility, user involvement in developing profiles, explicit plans for evaluation, and fairness to groups of providers. Then we presented the profiling system that we developed using Medicare data. This claims data set is well suited for profiling focused on quality of care because it includes information on the vast majority of persons age 65 and older, because people remain continuously enrolled, and because many services are covered (with the exception of many preventive services and prescription drugs).

The system presented here can also serve as a model for physician profiling in private managed care organizations if the limitations in many private health insurance claims data sets can be overcome. These data sets are often problematic because people change insurance plans, different sets of services are covered by different plans and (with the important exception of physicians who are members of PPOs or HMOs) physicians' practices are fragmented among many payers (Garnick et al., 1993, McNeil et al., 1992).

Until now, the measurement of quality in primary care has been underdeveloped compared with measuring hospital-based quality (Franks et al., 1993), a discrepancy that stems in part from the paucity of databases to use for studying the quality of outpatient care. However, a number of converging forces make the time appropriate to begin developing methods to evaluate primary care. First, newly available data sets such as the National Claims History file make it possible to track the care received in all settings by large numbers of patients and to link that care back to specific providers. Second, an explosion of activity in the development of practice guidelines gives us benchmarks that have been developed by clinicians to use to compare with actual practice. Third, as more medical care has shifted to the ambulatory setting and as more people gain access to health care through recent reform proposals, the interest in measuring the quality of primary care has increased (Daley et al., 1988, Palmer 1988). We believe that the physician profiling tool is a first step in developing a comprehensive system to measure the quality of ambulatory care in the United States.

Chapter 5

Developing a Quality Improvement Database Using Health Insurance Data: A User's Guide with Application to Medicare's National Claims History File

Abstract

As pressure continues to control health care expenditures into the 1990s, greater attention will likely focus on the effectiveness and quality of health care services. To support increasing demands for quality of care information, databases need to be developed that are comprehensive, relatively inexpensive to obtain for analysis, and flexible enough to focus on population, patient and provider denominators. Carefully used, claims data possess many of these desirable properties.

The DEMPAQ research team had at its disposal two years of claims for Medicare beneficiaries residing in Alabama, Iowa and Maryland to explore the complementary use of claims and chart derived information for quality of care analyses. Our work builds upon previous research on claims data and introduces two new results. First, we outline a strategy to clean a large administrative database from scratch using Medicare's National Claims History (NCH) File as an example. Second, we introduce several claims-derived technical innovations to support current and future quality improvement initiatives. This chapter will of particular interest to any policy analyst or researcher wishing to use a large claims database for quality of care assessment. In addition, an extensive discussion of possible database development strategies are presented to aid the technology transfer of our claims-based research and innovations to effective use by Peer Review Organizations for quality improvement initiatives.

Introduction

The value of the careful use of administrative claims data for health policy research has been consistently demonstrated in numerous research projects. Public and private agencies are increasingly turning to administrative data to provide a broad cost-effective overview of cost, utilization and quality trends in health care markets (Newhouse, 1991). One of the richest sources of inpatient cost and utilization data has historically been the Health Care Financing Administration's Medicare claims database. Recently, HCFA has successfully created a database with all ambulatory and inpatient claims called the National Claims History File (NCH). Medicare data are a huge resource only now beginning to be tapped for medical effectiveness, cost and quality analyses. The Medicare data provide valuable lessons to using future data sets from managed care organizations. Therefore, a thorough understanding of the Medicare data is crucial for analysts to use the data and for policymakers who need to take action based on those analyses.

In this chapter we summarize the key methodologies used in the DEMPAQ project to develop a patient population-oriented database derived from insurance transaction data. After providing a background discussion of the structure and uses of claims, we outline our strategy to clean, improve and construct a large database from scratch, using Medicare's raw National Claims History file. Next, we introduce several claims-derived technical innovations to support current and future quality improvement initiatives. This chapter will be relevant not only to those who wish to apply Medicare data for Quality Improvement (QI) analyses, but also people who are interested in the general application of private and public sector health insurance claims for the purpose of health service analyses.

Claims Data for Quality of Care Analyses

Many studies have successfully utilized administrative databases with encouraging results. Claims data have been used primarily as a tool for exploratory data analysis to establish sample populations and describe broad utilization and cost trends (Anderson, et al., 1990; Fisher, et al., 1990). In addition, many federally financed Patient Outcome Research Teams (PORTs) have included in their scope of work application of claims data for outcomes research (Lave, et al., 1993). Using claims, it is possible to change the focus from region, to provider, to patient at will. However, as other researchers have pointed out, claims data are not without their faults. These data are intended for paying bills, not monitoring episodes of care. The accuracy of a data item (e.g., a procedure or diagnosis code) may depend on whether it is needed for payment. The technical and methodological challenges are significant when claims data are being pushed to their limits of defining a clinical episode (McDonald and Hui, 1991).

While claims data may not have the clinical depth to conclusively determine whether the procedure performed on a patient was appropriate and effective, the efficiency of using large administrative databases for medical outcomes research remains one of the most convincing arguments for expanding their use and refining methods to accurately determine health outcomes. By "cleansing" claims data of inaccurate information and unneeded administrative transactions, a new information system can be harnessed to monitor changes in the financing, organization and development of health services. A claims-based Management Information System (MIS) can be used to monitor utilization and cost patterns as well as develop new risk adjustment models. Claims can be used as a "decision support system" for provider quality evaluation. Physician or facility-specific quality and utilization profiles can be developed to approximate a measure of cost effectiveness to help support quality as well as price competition among medical services provided to patients.

Contents of the National Claims History File

The Health Care Financing Administration administers the largest single insurance claims database in the United States. Claims data for approximately 33 million Medicare beneficiaries[1] are processed to reimburse providers and stored for later analysis (*Medicare and Medicaid Data Book*, 1988). Hospital admissions, physician office services, outpatient services, nursing home and home health services are all captured in HCFA's databases. Begun in 1989 and unveiled in 1992, HCFA's National Claims History (NCH) File project resulted in streamlining the processing of claims and the creation of a database that links the different types of claims together based on a common beneficiary identification (ID) number. By creating a unified claims database system, HCFA has produced a resource for extensive quality of care analyses (McCall, personal communication, 1991).

Part A and Part B Medicare Benefits Versus Institutional and Noninstitutional Claims

Part A and Part B are terms confined to the Medicare benefits package. Part A benefits include inpatient, hospice and nursing home. Part B benefits include outpatient services provided by physician and outpatient departments in hospitals. A noteworthy aspect about the structure of Medicare claims is the manner in which the files are organized. Medicare claims data are not structured by benefit package. HCFA pays claims to providers based upon who submits the claim, not the type of service performed. As a result, HCFA classifies medical providers seeking reimbursement into two separate categories: institutional and noninstitutional providers. An institutional provider is a facility such as a hospital, nursing home, hospice, or home health service. Noninstitutional providers are primarily physicians, group practices, and freestanding laboratory services.

To distinguish between Part A and Part B services, the data file has a "header" flag on each observation that will point to which of the Part A or Part B "trailers" one should refer to for analysis. The easier of the two databases to work with, the noninstitutional file, also has header and trailer portions for each claim. However, the noninstitutional file contains only Part B services and is mostly made up of physician claims.

---

[1] Medicare beneficiaries include Americans over the age of 65, the disabled population, and people receiving treatment for end stage renal disease (ESRD).

As it relates to quality improvement, there are two key features of the NCH. First, using the NCH, one can link Part B and Part A claims by patient or provider through a beneficiary identification number or a unique provider identification number. A more detailed account of our record linkage approach is presented in Appendix 1. Second, 100% of all claims are maintained on the file. As a result, a beneficiary who spends half the year in Phoenix and half the year in Minnesota has a complete record of all their reimbursed health services recorded at HCFA despite the two different sets of providers and data processors.

Advantages of the National Claims History Files over Past Medicare Databases
Significant improvements distinguish the NCH from other Medicare claims databases. The obvious improvement is the availability of 100% of Part B claims for 100% of HCFA beneficiaries. Previously, HCFA's so-called BMAD file contained a sample of 5% of all beneficiaries with 100% of their claims. Using a unique beneficiary ID, it is possible to describe the claims associated with a beneficiary no matter where the beneficiary resides (e.g., Vermont) and where they might seek service (e.g., Florida). The scope of HCFA claims for a beneficiary is somewhat limited to the carrier servicing a resident's claims. Another significant improvement in the NCH is the increase in edit-checked data fields. For example, ICD-9 diagnosis code fields are evaluated for completion as well as accuracy. Other edit-checked critical data fields for quality of care analyses include procedure code and Unique Physician Identification Number (UPIN). The lag period between date of service and date recorded on the NCH is significantly shorter than before. Approximately 99% of claims are available for analysis three months following the date of service. Finally, HCFA can array multiyear claims data with a common beneficiary identification number across time (Warren, et al., 1990).

An Inventory of the National Claims History Files
Although the National Claims History File title implies only a single file, the NCH is actually a "family" of three databases linked by beneficiary ID, as indicated in Table 1 (all tables are contained in Appendix 2). A description of these three databases follows with highlights of key issues relevant to using the data for quality of care analyses.

Health Insurance Skeleton Eligibility Write-off File (HISKEW): The HISKEW is the core database linking Part A and Part B claims by beneficiary ID. This file contains information on a beneficiary's age, gender, residence (in terms of state, county, and zip code), marital status and enrollment information. Socioeconomic status can be partially inferred by participation in a Supplemental Security Insurance (SSI) data field available on the file. Enrollment dates describe when a person started to use Medicare services with a specific beneficiary ID. Beneficiary terminations from the Medicare program, due to moving and death, are recorded as well on the "inactive file;" a separate but identically structured "active" version of the HISKEW file describing current Medicare recipients. It is important to note that a beneficiary may have more than one ID in a lifetime. For example, assume a husband and wife share the same base portion of their beneficiary ID, the husband's social security number. To distinguish between the husband and the wife, each has a Beneficiary Identification Code (BIC) describing their respective marital statuses. His BIC, A1, shows that he is the principal contract holder. Her BIC, B1, indicates that she is the spouse of the principal contract holder. Each person's BIC augments the principal contract holder's social security number, making each beneficiary uniquely identified. If, after ten years in Medicare, the husband dies, his widow's beneficiary ID would change in two major ways. The widow's social security number would replace her husband's number as the base portion of her beneficiary ID. Also, her BIC would change to indicate that she is widowed. As a result, the widow will have two separate beneficiary IDs and associated claims using both IDs. Strategies employed by the DEMPAQ research team to define a unique nonchanging beneficiary ID are later defined in the data cleaning section of the paper.

Institutional Claims Files: This file contains claims data primarily on hospital, nursing home, hospice, and home care center services. Outpatient procedures provided by a hospital outpatient department and ambulatory surgical center will be found on this database. In its raw form, the file has variable length records. These records contain data on the type of procedure, the diagnosis of the patient, the setting of the procedure, and the amount billed and reimbursed for the procedures. Each claim record contains information on the provider of service, the beneficiary ID and date when the service was performed.

Noninstitutional Claims Files: This file contains data primarily on physician services provided to beneficiaries. Another important source of noninstitutional claims is freestanding laboratories. Procedure, service date, diagnosis, place of service, provider ID and cost data elements are the most often used components of the file. It is important to note that physicians provide services in many different settings: a physician's office, inpatient, outpatient, nursing homes, a patient's home, etc. As long as the noninstitutional provider (either solo or group) is billing for a procedure, it doesn't matter where the procedure actually occurred. For example, at one point we noticed the second largest state (after their home state) rendering treatment to Maryland beneficiaries was South Carolina. More intriguing, most all of the services occurred in nursing homes. The research team's initial explanations ranged from entrepreneurial to morbid. After two months of analysis, we determined that the claims were for laboratory services provided for Maryland nursing home patients but performed by a South Carolina laboratory.

The Value of Carrier-Based Provider Files

Physician characteristics can be linked to the analytic files by provider ID. Under contract to HCFA, each of the Medicare carriers must collect information on physicians describing their specialty, credentials, board certification, medical school and practice location. From a health services research perspective, these variables can be used to determine if physician characteristics have any meaningful correlation to practice patterns. During the DEMPAQ project, we found these files essential in the determination of whether the provider ID recorded on a claim is associated with a physician's solo or group practice. Using the UPIN, the employer identification number (EIN), and designation whether a particular record was associated with a group or solo practice, we could distinguish a provider's group and solo practice IDs. It should be noted that without the carrier-provided provider characteristic file, or substantial revisions to the NCH, group practices cannot be uniquely identified.

Beneficiary and Provider IDs: The Keys to Developing an Episode of Care

Developing episodes of care allows an analyst to evaluate quality of medical services received in units relevant to the practice of medicine. Episode definition begins by defining a key event: the onset of disease or use of a major medical procedure (Hornbrook, 1985). All of the claims associated with the key event are then aggregated to develop a patient-specific episode summary. Linking claims records by beneficiary and provider ID is critical in episode development. Both the institutional and noninstitutional files contain the following information: specific procedures provided to the patient, dates of service, reimbursement amounts submitted and paid, the provider of service, place of service, and the diagnosis associated with the provision of the procedure. All claims can be linked to a beneficiary by a common ID. Physicians, facilities and other providers can also be identified by using a provider ID. Unfortunately, at this writing, each provider can use more than one ID (e.g., one for his/her solo practice, one for each group practice he/she participates in). Consequently, the unique identification of the claims associated with one unique provider ranges from burdensome to impossible. DEMPAQ's interim strategy used to remedy this problem, as well as HCFA's long-term solution, are discussed in more detail later as a database cleaning issue.

The following example provides a sense of the number and scope of raw claims transactions needed to construct an episode of illness record from institutional and noninstitutional files. An episode record can be constructed for a diabetic who recently had his foot amputated. The episode's focal point is the amputation admission. The episode begins with the most recent primary care physician consult before the amputation and ends with the consult and follow-up visits provided three months following the surgery. The amputation procedure can be identified by searching for an amputation admission using institutional claims data. The outpatient consults and follow-up visits can be recorded from the noninstitutional file.

Physician visits and consults to the patient while in the hospital were found on the noninstitutional file. Home care services received immediately following the amputation can be recorded on the institutional file.

Identifying Diagnosis and Procedure Information

Diagnosis information is recorded using five-digit International Classification of Disease Codes, 9th Revision (ICD-9). Outpatient procedures are described by the HCFA Common Procedure Coding system (HCPC) procedure code set developed by HCFA (which is a modified version of the American Medical Association Current Procedural Terminology {CPT} System). Inpatient procedures can be described by Diagnosis Related Groupings (DRG) or HCPC. If a facility bills for an inpatient procedure, a DRG describes the treatment. If a physician is billing for professional services accompanying an inpatient procedure, an HCPC describes the service provided. Any surgical inpatient procedure is also described by ICD-9 surgical procedure codes. In situations where a new procedure is being used and an associated HCPC code is not available, "local" HCPC codes are usually used. Local codes are maintained by state or region-specific carriers (who process claims on behalf of Medicare and are documented by HCFA). During our study period, a large portion of outpatient institutional records have no HCPC associated with the service provided. In our subsequent analysis, we recorded the charges and service use of these unnamed procedures but were not able to determine their description. HCFA has corrected this problem as of April 1, 1991 (personal communication with HCFA staff).

"Cleaning" Claims Files

All of the aforementioned features of the NCH have reduced the intensity of data cleaning activities, and it should be noted that HCFA is constantly updating the quality of the NCH. However, raw claims data are rarely ready to use "as is" for most analyses. We outline a comprehensive database cleaning strategy based on previous experience and our experience with the new NCH files. The goal of cleaning a claims database is the removal of any extraneous and inaccurate information that may be unneeded for a given analysis. For our project, data cleaning fell into two tasks: selecting a set of patients and preparing the claims for analysis.

Data Cleaning Task 1: Selecting the Population of Interest

The beneficiary ID is the common link of the individual NCH components. Once a beneficiary is identified, all of his or her associated claims can be selected. The HISKEW file contains the beneficiary information for all Medicare recipients including disabled persons, end stage renal disease patients, as well as the over age 65 population. The DEMPAQ analysis focused only on beneficiaries who belonged to the age 65+ Medicare program. We discarded from our analysis any end stage renal disease (ESRD) patients and disabled persons under the age of 65 years. Disabled persons over the age of 65 are automatically transferred to the standard Medicare contract at their sixty-fifth birthday.

For the DEMPAQ effort, HCFA provided beneficiary information for residents of Alabama, Iowa and Maryland. We chose to test our data cleaning strategy on the Maryland data due to our greater familiarity with the health systems in that state. Our study population included Maryland residents who were living in the state at any time during the year. As a result, a beneficiary who died part way through the year had their record flagged and kept for later analysis. To determine whether a beneficiary was a continuous resident or died within the year of observation, we used two snapshots of the beneficiary file. The first snapshot of the beneficiary file was taken at the start of the study period, July 1, 1990. The second snapshot of the beneficiary file was taken at the conclusion of the study period June 30, 1991. Using the participation and termination dates from the two points in time, we could identify continuously covered Medicare recipients and beneficiaries who died within the study period.

When selecting beneficiary IDs, it is necessary to "BIC-equate" the IDs. Recall that a person can have changing IDs during a lifetime. BIC-equating is a process by which the person's multiple IDs across time are recoded. To provide this feature, HCFA will change a person's beneficiary ID registered on a claim if it differs from the original ID at the start of a period. A cruder, yet more straightforward, process to remedy changing IDs is to match claim and beneficiary information based on a combination of birth date, gender, and social security number (SSN). These data fields are available on the HISKEW as well as the claims database. The goal of BIC-equating or birth/gender/SSN matching is the identification of a unique beneficiary ID to track a patient over the course of one or several periods. This is not a trivial issue; in a one-year period, we found that approximately 16% of the total claims could be mismatched due to changing ID numbers.

Data Cleaning Task 2: Preparing the Claims Data for Analysis

The cleaning phase had three objectives: all of the files were first converted into a form useful for rapid extraction of fields and records; second, adjudications or transactions that were claims not associated with service delivery were removed; and finally, the claims of the study population were selected. To achieve these objectives, we completed the following steps:

- *Convert Claims to a Fixed Format:* Each of the NCH's raw claims records has two components: a header and a trailer. The header portion of the claim record contains the claim ID, beneficiary ID, provider ID, principal diagnosis codes, and basic Medicare contract information. The trailer portion contains information on the procedures performed, the specific diagnosis code associated with the procedure billed, charges submitted, final reimbursement, type of service and other information. Noninstitutional file records can have up to 15 trailers per header. In addition, the trailers can have variable record lengths (i.e., it is possible that each record may be of a different length). Variable length records are inefficient for health services analysis. Therefore, variable sized records were converted to a fixed length. The header and trailer were combined to form a single record describing the service provided. This level of detail is commonly called line-item detail. In contrast to claim-level detail, line-item units of observation reflect individual services provided on a given date, but submitted together on a single claim form. As a result, several items may constitute one claim record.

  Table 2 presents a state-specific summary of claims, line items, services and costs with the study population. The differences between claims- and line-item detail described in the data cleaning steps is evident. A further disaggregation of utilization information is service-level information. Service information is contained in the line-item detail of a record as a variable indicating the frequency a specific procedure code listed on the record was performed. For example, a line-item states a chest x-ray was performed on a given date of service. The service frequency field indicates how many unique chest x-rays were performed on a given day. It is important to note that the service frequency field has been used to describe anesthesiologist services in terms of minutes of an anesthesiologist's time. As seen in Table 2, Alabama and Maryland were found to have roughly twice the number of services in comparison to Iowa. Upon closer examination, we found the problem in Alabama and Maryland to be largely attributable to anesthesiology services. Without identification of the anesthesiology anomaly, the number of services can be significantly over-counted.

- *Select Claims for the Study Population*: Once the claims are "fixed," the beneficiary ID listed on the claim is matched to the beneficiaries selected from the HISKEW file. This process reduces the number of line items used to only those individuals included on our study. It is critical that BIC-equating, described in the HISKEW beneficiary selection step, is completed before this match is made. Beneficiaries selected from the HISKEW file with no claims history will be discarded from the sample population. As a result, the DEMPAQ project analyzed only health service users.

- *Identify Adjudications:* Claims databases are the by-product of our fee-for-service reimbursement system. One common feature of claims processing is the adjudication process. For example, take the following sequence of events: 1) A physician submits a claim and a record is created; 2) HCFA decides that the procedure could have been substituted by a less expensive treatment; 3) HCFA creates an additional claims record that is identical to the first, except that the payment is the negative amount of the original (this action "zeros" the amount paid); 4) HCFA creates a final record that is identical to the first, except payment reflects the cost of the less expensive treatment. The net result of this adjudication is three claims, representing one service performed. For the purpose of analysis, only one service should be recorded. We found it best to use claims replaced units when this problem occurs. In any analysis with claims data, this step should be a necessary precaution to avoid double or triple counting of claims. HCFA has since made available (as of August 1993) only final adjudication records to researchers.

- *Out of Scope Claims:* Once the claims were fixed and adjudication information was dropped from the final claims files, we deleted out-of-scope claims. For this study, we were not interested in skilled nursing facility (SNF), hospice, and home care claims, so these records were dropped from the analysis. Also, persons who received care for part of the year from HMOs were deleted from the analysis, given that claims dates are unavailable from HMOs, which are capitated.

Creating Analytic Files

Although cleaning the claims files goes a long way toward preparing the data for research, they are still not ready to apply to most quality improvement analyses. Logical units of service used in quality of care analyses (e.g., a visit, a person, or an episode) are usually made up of many transactions. Data elements from other databases (e.g., beneficiary and provider files) must be combined with claims records for many analyses. The goal of creating analytic files is to provide a database that can be readily used for research and management analyses. Using cleaned data records, we developed several analytic files described below.

Claims Extracts

From the cleaned noninstitutional and institutional claims files we created significantly stripped-down versions for potential use on a PC platform. The institutional file was split into two so-called "extracts": a Part A institutional file and a Part B institutional file. The Part A institutional file contained line-item information regarding the study population's hospital admissions. The Part B institutional file contains information on services delivered to the beneficiary in outpatient settings, but provided by an institution (e.g., a hospital outpatient department).

Developing an Enhanced Beneficiary File

The development of person-specific analytic files was one of the objectives of the DEMPAQ project. The HISKEW served as our base. To this, we added additional information describing the person's "functional" status, source of primary care and case-mix category. Functional status was inferred from flags indicating whether the beneficiary had died or was institutionalized within one year. The institutionalized flag was derived from claims data and was based on the length of time associated with three months of contiguous services in a nursing home or psychiatric facility.

An objective of the DEMPAQ project was the evaluation of using ambulatory claims data for case-mix adjustment. The Johns Hopkins Ambulatory Care Group (ACG) case-mix system was used to provide a beneficiary specific case-mix category that infers illness burden in terms of severity and duration of a medical condition (Weiner, et al., 1991). ACGs were used as the primary method to case-mix adjust expected service utilization and cost described in physician specific profiles. The ACGs constitute a constellation of 52 unique patient categories. Comorbidity is also explicitly taken into account using Ambulatory Diagnostic Groups (ADGs), a byproduct of the ACG case-mix assignment software. All of the beneficiaries in the study population were assigned an ADG based on their ICD-9 diagnosis information provided on the claims files. The ACG assignment process also provided an indirect validity check of valid ICD-9 codes. If the ICD-9 codes listed on the claims did not exist, the ACG grouper would report the level of mismatch. We found the level of mismatched diagnosis codes in the state of Maryland to be below 5%.

Each beneficiary's provider contacts were evaluated to find evidence of repeated visits to primary care providers. Based on this analysis, we developed an algorithm to identify a beneficiary's primary care source (PCS). A PCS is defined as the primary care physician, specializing in general practice, family practice, internal medicine, or osteopathic medicine, who has seen a particular beneficiary most frequently. Encounters by primary care physicians occur in an office setting and constitute a "face-to-face" encounter in which the physician was likely to spend a significant portion of time with the patient to ascertain his or her condition. Face-to-face encounters were distinguished by a set of HCPC codes and the place of service reported on noninstitutional claims. Of the primary care physician face-to-face contacts associated with a beneficiary, the provider with the most contacts was designated as the PCS. Ties between competing providers were broken by charge information.

A primary care source was further characterized as a primary care physician who had at least 25 patients. To ensure that the physician was continuously practicing medicine in-state during the year, we further restricted the PCS designation to providers with at least two visits made by any of their primary patients during a quarter of the year. As a result, the enhanced beneficiary file includes a beneficiary's PCS and relevant provider information, such as specialty. Over 75% of all study populations' beneficiaries could be assigned a primary care source. The remaining study population did not have enough general medicine encounters to justify assignment to a primary care source (Parente, et al., 1992).[2]

Table 3 describes the breakdown of beneficiaries included in the study population from any Medicare-eligible state resident to those who were assigned a PCS with 25 or more patients and continuous provision of service. As expected, the study population gets smaller from each cut of the population. However, the PCS-assigned population remaining is still the majority of all beneficiaries receiving services in each state and can be assumed to be a fairly accurate depiction of average health service users in each state. On average, at least 70% of the study population had a PCS.

In terms of sociodemographics and office practice characteristics, we found Alabama to be the state that best represents the entire study population. As a result, the Alabama population is used as a proxy for the entire study population as more detailed tables are presented.

Table 4 provides a comparison of the case-mix of the Alabama total beneficiary population assigned and not assigned a PCS. Case-mix strata include age categories, gender, and the number of comorbidities associated with each patient. The number of comorbidities is based on the number of unique diagnostic groups (i.e., ADGs) produced from the ACG case-mix classification system. The patients with an assigned PCS are younger and tend to have a greater proportion of males than the assigned PCS population. The pronounced difference between the two populations is the number of comorbidities. The non-PCS population is clearly healthier. Much of this result is attributable to the fact that there are roughly 50,000 nonusers in the non-PCS category. Although, one may argue that nonusers in Medicare may be actual self-pay patients, the percentage of the Medicare population paying their own way has to be fairly small if beneficiaries are rational. Even discounting the effect of including nonusers, the non-PCS assigned population is still significantly healthier than that assigned a PCS. Since the focus of DEMPAQ was methods to evaluate quality of care, we feel the focus on the population is justified by their more intense experience with the health care system.

---

[2] Our decision to organize our analysis around the PCS was motivated by a desire to establish a medical practice denominator to which a patient is assigned. We also wanted to test the logistic plausibility of Medicare patients adopting a managed care "gatekeeper" approach by examining their existing patterns of care. Those wishing to use only the provider ID used on the claim should beware of problems identifying a unique physician or practice discussed in the text.

Another reason for incorporating case-mix adjustment into the beneficiary files was adjustment of physician-specific total resource units. That is, if you take an average of all of the cost of the services a PCS's patients incurred with any physician or hospital and compared it to the state average, one might say a PCS is billing more or less than the state average. In Table 5, we display a randomly selected PCS's cost per patient by several categories to demonstrate the policy implications of using unadjusted versus case-mix adjusted average costs to rate the cost effectiveness of a provider in a gross fashion. For ambulatory, hospital and total charges, the physician would be considered over average if compared to his peers based on a state average. However, the ACG-derived case-mix adjusted cost per patient shows the physician to be billing less than average in all services displayed except laboratory testing. It should be noted that using the Provider ID from the claim would provide nonexclusive matches between physician and patient. As a result, case-mix adjustment would be far more complex due to accounting for each patient's individual contribution to a physician's total cost. PCS assignment provides a simplification to this problem and a managed care application germane to national Medicare policy development.

Primary Care Source (PCS) Quality and Utilization Profile Files
Using the extract claims and enhanced beneficiary files, we created physician-specific profiles. These profiles represent provider-specific summaries of a patient's quality of care and utilization measures. Two types of profiles were generated for the DEMPAQ project: condition-specific profiles and office practice profiles. The condition-specific profiles described process and outcome measures associated with a PCS's patients for a given medical condition. The conditions analyzed included diabetes, chronic obstructive pulmonary disease (COPD), congestive heart failure, osteoarthritis, hypertension and ischemic heart disease. The office practice profile focused on demographic, utilization and cost statistics associated with a primary care source's patients (Lawthers, et al., 1993; Garnick, et al., 1993). Inpatient, as well as ambulatory, services were linked on a beneficiary level and aggregated to the provider level to create the profiles. Examples of selected tables from the quality and utilization profiles are presented in Garnick, et al., 1993.

The generation of physician profiles involved a multistep process using a relational database approach. Except for one of the tables, we were nearly always able to create the profiles on a fast PC with a large hard drive (i.e., 1.4 gigabytes). Our only limitation using a PC was the longer time requirements to create the profiles in comparison to a mainframe. The profile generation process can be summarized in five steps. The condition–specific profile for diabetes is used as an example.[3]

Raw Claims to a Diabetes Profile:

1. Identify beneficiaries with one or more "face to face" encounters with a primary diagnosis of diabetes (i.e., ICD-9 250.XX)
2. Select all claims for these beneficiaries.
3. Aggregate the claims for each patient. Develop count fields of the relevant services.
4. Assign a primary care source.
5. Summarize count fields by primary care source and record on an analytic file.

Technical Innovations

To create the physician profiles, two innovative strategies were used to extend the abilities of the NCH database: Development of a resource unit table to uniformly price all services and generation of a solo and group practice identification numbers.

---

[3] For more information on the profiles generated for the DEMPAQ Project, consult Garnick, et. al., 1993.

Resource Unit Table

There were three different types of procedures identified in the DEMPAQ claims analysis. First, we used DRGs used to describe inpatient procedures.[4] The second type of procedure involved services billed using HCFA's procedure code system of HCPCs. The third type of procedure code used was of the nonexistent form. Missing procedure codes were concentrated in the institutional Part B claims. These records were associated with outpatient services provided in a facility rather than a provider's office. These claims were paid and have diagnosis information, but do not contain any consistent indication of what the procedure was. A proxy measure for a small percentage of these missing procedures is HCFA's revenue center code that describes, in the broadest terms (e.g., radiology), what procedure was performed.

In an attempt to minimize the bias of any region-specific service pricing and reimbursement policies, we decided to develop a standard reimbursement for all of the types of procedures used in the claims data. For inpatient information, we based our standardized reimbursement on the average amount paid by HCFA per DRG in the state of Maryland (the "Maryland standard" was used across all three states). For outpatient procedures, the RBRVS weights for each HCPC, ignoring geographic factor price differentials, were multiplied by the approximate 1992 fact or $31 to produce standardized reimbursements (*Federal Register*, June 5, 1991).

In those situations in which no RBRVS HCPC was available, but a procedure code was listed on the claim, the Maryland average reimbursement for the procedure was used. The thorniest issue of all was the pricing of the nonprocedures. We made the decision that their cost and utilization were necessary to create an accurate analysis. Since the nonprocedures could not be priced on a line-item basis, they were never associated with a fee, although they were treated as one giant procedure in the final summary of utilization and cost of a primary care source's patients. We later learned from HCFA that the nonprocedure problem was confined to medical office services provided during our study period. The problem was resolved by HCFA on April 1, 1991, in preparation for analyses leading to the development of outpatient DRGs.

Solo and Group Practice Identification

Our second innovation was the creation of unique identifiers for solo and group practices. We found many physicians practicing in group practices as well as their own solo practice in all three states. The first step to identifying any practice was to test whether the Provider ID field on the claim was unique to one physician or a group practice. To complete this test, we matched the Provider IDs between the claims and the carrier-supplied provider files. As Table 6 shows, we found the number of unique provider IDs (13,229) used in Alabama noninstitutional claims to be roughly two times larger than the actual number of physicians indicated by the number of UPINs (6,954). Using our strategy of matching provider IDs and looking on the carrier-supplied provider files for group practice identification, we were able to develop a set of recoded Provider IDs. We found our method of recoding provider IDs indicated that roughly a third (2,722 of 7,135) of the unique IDs used on the claims were associated with solo practices.

Table 6 also highlights the dangers of relying entirely on UPIN to identify a group practice. UPIN was a required field for only half of our study period but its absence from 83% of patients claims motivated our decision to develop our own method to identify solo and group practices. With UPIN now a required field, some of this problem will be eliminated. However, without explicit identification of whether the claim is submitted by a group or solo practice, our method remains a burdensome alternative to correct group practice identification.[5]

---

[4] ICD-9 surgical procedures codes were also found on the inpatient claims. We found the consistency of the ICD-9 codes to be lacking and chose to favor using DRGs exclusively to describe inpatient procedures billed by a facility.

[5] Without the use of the carrier-supplied provider file, practice type identification would have been impossible.

Our identification consisted of matching all possible provider ID numbers to the carrier-supplied provider files for each state[6] and then distinguishing group practices based on employer identification number (EIN) and group practice indicator. The EIN is the common tax identification number used by a group practice to report its income to the Internal Revenue Service.

We found that specialty information from the claims and the provider file can be further used to distinguish between different types of group practices. As Table 7 shows, we could distinguish three types of group practices. The first type is a group practice consisting of only primary care physicians. The next two types have a mix of primary care and specialty physicians. When more than 50% of the physicians are primary care providers, the group practice is denoted as majority primary care group practice PCS. It should be noted the difference in number of physicians listed in Table 7 for Alabama is significantly smaller than in Table 6 due to a focus on primary care sources. Once the 25+ patients and greater than two visits a quarter rule is applied to the Alabama PCS providers, the number of practices is further cut from 1,370 to 866.

Accuracy and Reliability of NCH Claims Data

Our experience with HCFA's NCH file provided a preview of some of the methodological and technical hurdles required to develop regional or national databases of patient encounter information based on claims data. One of our major underlying concerns through the project was the validity of the data, particularly diagnostic and procedural fields. We were impressed to find that the quality of the diagnostic information was not as bad as many have thought, particularly physicians. Two separate analyses highlight the reasonable quality of the ICD-9 codes used in the analysis.

The first of the two diagnostic accuracy analyses was a by-product of the case-mix adjustment process. The ACG grouper was unable to classify less than 5% of the diagnosis codes it read. The ungroupable codes were nonexistent, local codes used by carriers, or garbage. To contend with the possible problem of zero-filling the right-hand side of the diagnosis codes, we devised an algorithm to work backwards from five to four to three-digit identification to group the diagnosis codes. The degree of zero filling did not produce appreciable differences in the case-mix distribution. The DEMPAQ project's formal test of the diagnostic accuracy of claims data was a comparison of chart and claims data reliability to detect a diagnosis or procedure.

Expected and Suggested Enhancements to the NCH File

While working with the NCH file, we have maintained active communication with HCFA regarding our findings and suggested improvements. The quality of the NCH file has dramatically improved over the course of the DEMPAQ project. Our understanding from HCFA is that our current NCH claims are now representative of an older generation of files. The new generation files contain: fixed formats; consistent edit-checking of all the critical fields used in our analysis; and a significantly smaller size due to a removal of claims processing information data elements.

Several further enhancements have been suggested to HCFA to improve the quality of the NCH database. These include but are not limited to:

- An edit-checked UPIN (against a national registry) on every claim;
- Collection of a group practice's EIN on the claim record as well as flag for solo versus group practice practitioner; and
- Use of a unique nonchanging patient identifier (e.g., SSN)

---

[6] Iowa and Alabama beneficiary claims are processed by Iowa Blue Cross/Blue Shield and Alabama Blue Cross/Blue Shield, respectively. Maryland's claims are processed by two carriers: Blue Shield of Pennsylvania (Prince George and Montgomery county beneficiaries) and Blue Cross/Blue Shield of Maryland (the rest of Maryland). As a result, a provider file from each of Maryland's carriers had to be obtained to identify group practices.

An edit-checked UPIN provides a means to uniquely identify any physician despite their group practice affiliations. Physician profiling could focus on either the individual physician level or the practice level. At the moment, changing the denominator from physician to practice remains a fairly complex task with some possibility for error. The UPIN is based upon a national system of unique physician identification. Matching UPIN to a national registry will minimize the chances of aggregating information up to the physician level inappropriately.

Collection of a group practice's EIN and a flag to distinguish a service performed by a provider wearing their group practice or solo practice hat would remove reliance from the carrier-supplied provider file for this level of analysis. Given the significant proportion of group practices acting as a patient's primary care source, it will be important to correctly identify their unique practice style. Providing researchers with the carrier-supplied provider files is an untenable option. Although HCFA provided general guidance regarding the critical fields of the provider files, none of the four files we encountered shared a common format.

Future longitudinal analyses using HCFA data will be dependent upon correct identification of a beneficiary across time. At present, HCFA continues to use the combination of the principal contract number and BIC code to uniquely identify a beneficiary. We suggest that HCFA simply use each person's social security number as the primary component of a beneficiary ID. A person can then be identified across time without regard to their marital status and other factors.

Technical Considerations to Using the National Claims History File

The HCFA NCH is the largest claims database system in the world. We found that using the database effectively was a function of the clarity of the research team's objectives. That is, with data files this big it is difficult to perform an exploratory data analysis, even with just a 5% sample. The DEMPAQ project proposed three major tasks to be completed with NCH data: generate physician profiles; query the database for frequencies and simple statistics; and complete complex ad-hoc data requests. Recognizing that our technical capabilities throughout the project were never static (i.e., within three years, our hard drives increased by an order of magnitude in size for roughly the same cost), we describe our perception of the technical requirements to use claims data for quality of care analyses.

Technical Requirements: Generating Physician Profiles

Development and Enhancement: The primary requirement for developing physician profiles is access to a mainframe computer to either select the data elements one would download to a PC platform or to actually crunch the data to create the profiles. For DEMPAQ, we generated profiles using a combination of both approaches. We envision mainframe utility programs that can be used to extract key data elements to be downloaded to a PC. Once on a PC, either programming/statistical software (e.g., SAS) or a relational database package (e.g., FOXPRO) could be used to generate the elements of a physician profile from the extracted claims, beneficiary and provider records. The complexity of generating a provider profile increases exponentially as case-mix adjustment and episode of care features are woven into the product.

Printing: Once the elements of physician profile are generated, the profile can be easily printed. Organizations wishing to develop a quick means of printing and sending a profile should first develop a common analytic file that can easily be based on a PC platform. From this file, a user interface can be designed to simply specify the name of the provider and what components of a physician's total practice should be printed. Ideally, we think CD-ROM technology would be suited for this task. In the case of Medicare peer review activities, HCFA could create the files internally or contract out the task of generating periodic physician profile summary files available for PRO use.

Technical Requirements: Query the NCH database for Frequencies and Simple Statistics

Many times a canned summary of physician or patient services is not sufficient to address a policy question affecting a population. A query system of the NCH would be useful to generate cross-tabs of key fields. With a mainframe, several canned query systems are available. To attempt developing a PC-based query device, one must have storage space into the gigabytes and a capable processor. Our experience indicates that a parallel processing technique or some other innovation to dramatically increase processing speed is needed to make this a viable option. Based on the rate of innovation in information systems developments, a PC-based query system would be a reasonable goal within a few years.

Technical Requirements: Complete Complex Ad-hoc Data Analyses
If a physician profiling or a simple query system is not a sufficient resource to study an issue, the remaining alternative is an ad hoc data request system. Given the size of the NCH files, (even on statewide level) an ad hoc approach cannot be reasonably completed on anything less than a mainframe. This assumes a person needs to run against the full set of claims every time to complete an analysis. For claims previously selected for a specialized population (e.g., coronary artery bypass patients) a PC-based approach to ad hoc analysis is reasonable but still fairly limiting in terms of time resources. For a detailed summary of some of the data issues arising from specialized database development in the PORT projects (consult Lave, et al., 1993).

Summary

The quality improvement process is an iterative process dependent on data that can show changes in the quality of a health care system. Berwick (1989) challenged researchers to develop common methods and health systems databases to provide a common ground for physicians, insurers and analysts to generate an informed continual conversation about methods to improve health care delivery. Jencks and Wilensky (1992) responded to the trend toward quality improvement on behalf of HCFA with the announcement of the Health Care Quality Improvement Program. The development of the NCH as an approachable comprehensive claims database for research and policy development stands as one of the products of this research. However, the existence of a database is only a prerequisite for meaningful analysis. Careful cleaning of any private or public administrative database is almost always a given (Garnick, et al., 1993). The subsequent development of denominators and specialized innovations to achieve the goals of a project is a path with many decision points and little guidance.

By sharing our experience with the NCH database we set out to describe a systematic approach to working with a large claims database. Three steps summarize our approach: clean; reduce; and enhance. We began by cleaning the database with a series of sensitivity tests and frequency counts. Our initial data cleaning strategy was not our last. Once we were reasonably confident of the quality of the data elements, we reduced the databases to only the file elements needed. Our goal was to minimize the time required to process the data as well as explore the possibility of downloading key elements of the extract files to the PC. Finally, as our research objectives focused clearly, we began to enhance the database with innovations and summarize key elements into analytic files.

Working with a file the size of NCH presented many unique opportunities to pursue region-wide analyses. However, the costs of simply creating the analytic files were fairly high. The irony of working with a large database is the vast range of research questions generated by data with nearly infinite possibilities and a very finite budget to just clean the data and make it available for a finite number of tasks. Although HCFA has significantly improved the quality of the NCH, the staff and computing resources to work with the file cannot be underestimated. One principle of the DEMPAQ project was to avoid building black boxes that a select group of programmers create which no one can open up and inventory all the parts. Iezzoni (1991) warns of the perils of black boxes that cannot be validated externally. Although the entire DEMPAQ research team was involved in all aspects of profile development, the final programming remained largely a black box due to logistical realities. We found the costs of maintaining a high degree of communication between researcher and programmer high yet absolutely critical. These costs can also never be underestimated if the process is to be continually updated.

Future changes in the delivery of health care may retire some of our approaches. For example, as HCFA moves to a "coordinated care" delivery system with gatekeepers, PCS assignment will be meaningless. However, in the interim period during the transition, PCS assignment could provide a reasonable proxy for strategies to increase quality of care through a gatekeeper system.

As health reform unfolds, the demand for the information regarding the quality and efficiency of health services will increase dramatically. Private and public insurance agencies are aggressively seeking ways to get the most out of their resources' health care dollars. To support this increased demand for information, comprehensive databases that focus on populations must be developed. Several researchers have already noted that "managed competition" could affect competition in quality, as well as price. Accurate measures of cost and quality are critical to assessing the cost effectiveness of a medical provider or treatment. Health insurance claims data hold great potential to meet many of the future demands of information-driven health care systems. Our experience illustrates the usefulness of the NCH and highlights the need to treat it with special care. The NCH has great potential and we welcome others to try out approaches to tame it.

Appendix 1

Appendix One

Linking Physician Inpatient Claims to Hospital Inpatient Claims

Many claims analyses require capturing the total inpatient costs per admission. Most claims systems, including those at HCFA and Maryland Medicaid, maintain hospital bills in one data base and physician bills, including the physician component of an inpatient stay, in another. In order to obtain a total cost per admission, it is necessary to link the physician bills to the appropriate admission. This can be problematic due to the sizes of the claims data bases and if, as is usually the case, the physician bills do not contain a reference, i.e., admission date, to the inpatient stay. The following is a description of an approach that was used to accomplish the linking of physician inpatient claims to corresponding hospital admission dates using complete Medicaid claims data from Maryland and again using HCFA claims data from Maryland, Iowa and Alabama.

One goal in designing this approach was to develop a method that could be easily duplicated by a mid-level research associate using the most common mainframe utility software and one of the major statistical software packages, e.g. SAS or SPSS. The task required that physician claims be matched to hospital claims to obtain pertinent admission information, e.g., admission date. The major impediment to this task was that the physician claims usually contain only a single date corresponding to the actual date that the physician provided service, whereas the hospital claims usually contain a beginning and ending date corresponding to the range of dates the patient was in the hospital. Most commercially available match software only allow for an exact match between two pieces of data. In this case, we need to match a single date to an implied range of dates.

The first step is to use a selection utility, e.g. SYNCSORT, to select record identification information from the physician claims: person ID, date of service and a record number. This extract record is usually much smaller in length than the original claim. As an example, with an 11 character ID, 6 character date and 9 character record number, 6 million claims, average for the three states mentioned above, would require 156MB of space (less than 1 tape), which is a small to medium size file for most mainframes. Furthermore, the size of the file would be reduced substantially more since the selection would be restricted to physician inpatient claims only. This file was sorted by person and date during the selection.

The second step in the process is to select pertinent information from the inpatient claims: patient ID, admission date and beginning and ending dates of service. Note that since most hospitals bill on a cycle and not ny admission, the beginning date of service will not necessarily be the admission date, but rather the first day of the billing cycle. Again, the resultant file is small by mainframe, and even some PCs', standards. The file should be sorted by person and beginning date of service.

The next step involves converting the selected inpatient records, one record per bill, to one record per day in the hospital. So, for example, if a record indicated 10/1/93 as the beginning date of service and 10/31/93 as the ending date of service, it would be converted to 31 separate records, each one containing the person ID, single date of service (which now corresponds to a single day in the hospital), and admission date. This task can be easily accomplished using SAS or SPSS. The resulting output record is smaller than the input record because the beginning and ending dates of service were replaced by a single date of service, however, based on an average length of stay of 12 days, the output file should be about 10 times as large as the input, which is still easily handled on a mainframe. As an example, in one study [DEMPAQ], we had 106,529 admissions for 64,257 admitted patients. The record ID was 11 characters and dates were 6. The average length of stay was 12 days. This results in an input file size of 3MB (LRECL=29 times RECNMR=106,529) and an output file size of 30MB (LRECL=23 times RECNMR=106,529*12).

Next, we use a commercially available match package to do a straight match of the selected physician claims to the "days in the hospital" file created in the previous step. The admission date is attached to the physician claims.

Finally, we re-sort the physician claims extract, which now has the admission date attached, by the claim record number. This file can then be matched back to the original claims on record number to attach the admission date to the original claims.

Following this process, we achieved a match rate of 93% for Medicare claims for Maryland in 1991, 87% for Alabama and 91% for Iowa. The non-matched physician claims are attributed to keypunch errors and missing inpatient claims (possibly due to adjudication).

Appendix 2

Table One - National Claims History Files

National Claims History Files	Sampling of the Contents of Files
Beneficiary File (HISKEW)	Beneficiary Information: Unique identification number, age, sex, beneficiaries state, county and zip code, marital status
Institutional Claims File *Contains Part A (inpatient) as well as Part B (outpatient) services.*	Claims Submitted by Institutional Providers (e.g., hospitals, nursing homes, hospices): Patient ID, provider ID, date of service, type of service, procedure and diagnosis.
Non-institutional Claims File *Contains only Part B (ambulatory) services.*	Claims Submitted by Non-institutional Providers (e.g., physicians): Patient ID, provider ID, date of service, type of service, procedure and diagnosis.

Table Two - Service Utilization and Cost Breakdown by State

Service Use & Cost Information per Beneficiary	State		
	Alabama	Iowa	Maryland
	Total Study Population	Total Study Population	Total Study Population
Total Claims	6,484,340	4,795,853	6,238,481
Total Line Items	12,853,752	9,076,118	11,657,476
Total Services	27,264,197	12,643,598	21,141,754
Total Amount Paid $$	2.6 billion	1.4 billion	1.9 billion

Table Three - Study Population Breakdown

Study Population Denominator	State		
	Alabama	Iowa	Maryland
Total Beneficiaries in State	474,332	384,625	447,145
Study Population (health service users)	415,723 (100%)	350,731 (100%)	389,765 (100%)
Study Population assigned a PCS	328,740 (79%)	277,727 (79%)	286,221 (73%)
Study Population assigned a PCS with >25 patients and >2 visits a quarter	298,394 (72%)	254,914 (73%)	264,885 (68%)

Table Four - Case-Mix of Patients with and without PCS Assignment

Case-Mix Measures	Study Populations by PCS Assignment		
	% of Patients with PCS	% of Patients without PCS	% of All Patients
Age Strata			
- Aged 65-74 years	56%	59%	57%
- Aged 75-84 years	35%	30%	34%
- Aged 85+ years	9%	11%	9%
Gender Strata			
- Male	36%	43%	38%
- Female	64%	57%	62%
Comorbidity Level			
- 0 Conditions	0%	58%	18%
- 1 Condition	10%	15%	11%
- 2-3 Conditions	30%	16%	26%
- 4-5 Conditions	28%	7%	21%
- 6-9 Conditions	27%	4%	20%
- 10+ Conditions	5%	0%	4%
Number of Beneficiaries	328,470	145,862	474,332

Table Five - A Physician's Actual Resource Units Compared to the State Average and Case-mix Adjusted Expected Costs per Patient

Resource Unit Category	Cost per Patient (actual to comparison (A/C) ratio)		
	Actual	State Average	Expected based on Case-Mix (ACG)
Ambulatory Services	1577.19	1493.23 (1.06)	1732.65 (0.91)
Laboratory Services	102.26	77.09 (1.33)	88.26 (1.16)
Hospital Services	2778.98	2419.29 (1.15)	2920.92 (0.95)
Total Services	4458.43	3989.61 (1.12)	4741.83 (0.94)

An A/C ratio greater than 1 indicates a higher actual cost than the comparison value.

Table Six - Results from Three Methods to Identify Unique Medical Practices Based on Alabama NCH Part B Claims

Unique Practice Identification Criteria	Provider ID Type for Alabama NCH Part B Claims		
	Submitted Provider ID	Submitted UPIN	Re-Coded Provider ID
# of Unique IDs	13,229	6,954	7,135
# of solo practice physicians with multiple IDs	1,498	0	0
% of line items of missing IDs	0	34% of line items	0
# of patient associated with missing IDs	0	346,511 83% of patients	0
Actual Practices - Solo - Group - Other/unknown	4,777 5,388 3,064	1,237 4,843 874	2,722 2,614 1,799

Table Seven - Primary Care Source Practice Type Breakdown

Provider Type	State		
	Alabama	Iowa	Maryland
Total PCSs	1,370 (100%)	1,194 (100%)	1,702 (100%)
Solo-practitioner PCSs	815 (59%)	900 (75%)	1,440 (85%)
Group Practice PCS	555 (41%)	294 (25%)	262 (15%)
Primary Care Group Practice PCS	490 (36%)	202 (17%)	165 (10%)
Majority Primary Care Group Practice PCS	65 (5%)	92 (8%)	97 (5%)

Chapter 9
Agreement Between Physicians' Office Records and Medicare Claims Data

Abstract

Methods
Office records that had been submitted for the record review (see Chapter 11) were abstracted for a limited set of items that appear in the claims. Data were collected using a computerized record review system that was specially designed for the reliability study. Abstractors had to look for a specific mention of a diagnosis or test and were blinded to the occurrence of any of the items in the patients' claim history. To compare the claims data with office record data, a translation table was created specifying the codes to be considered matches in the claims files for the items searched in the office record.

Data were analyzed by calculating percent agreement between claim and record and a kappa statistic.

Results
The office record and the HISKEW file agree well for name and gender only. Patient zip code, date of birth, and death do not match well.

For diagnoses, the best agreement (as measured by kappa) was for the most serious major chronic diseases. An unblinded rereview of 21 mismatched cases allowing clinical inferences show much better agreement.

For lab tests and procedures, agreement between claims and records was worse than for diagnoses. Not unexpectedly, tests and procedures appeared more frequently in the claims than in the record. These results suggest that much testing is being conducted that the primary care physician may not be aware of.

Discussion
Claims tends to be highly specific, but not sensitive. The usefulness of claims depends on the nature of the question being asked.

Introduction

Measures of quality are the linchpin of successful health care reform for providers and payers. In the 1990s health care payers are eager to provide their customers with access to the most cost-effective health care systems. To reach this goal, they must be able to compare costs and performance of competing systems. To be legitimate, these comparisons require accurate measurement. Providers need to be able to measure their own performance as part of the process for improving it. The measurement required to achieve these objectives emphasizes the need for accurate data.

Two data sources frequently used to make these measurements are medical records and administrative billing sources, or claims. While the development of DRGs triggered extensive investigation of the accuracy of hospital medical records (Roos 1989, Hsia 1988, Lloyd 1985, Feigl 1988, Iezzoni 1988) and use of hospital billing and discharge data (Wennberg 1987, Roos 1985, Roos 1990, Luft 1986, Anderson 1990, Iezzoni 1992, Maronde 1989), less is known about ambulatory medical records (Palmer 1988, Horner 1991, Ellis 1990, Lurie 1992). Hospital medical records have been the *locus* of extensive quality assurance activities and billing verification. In fact, hospital records have become the gold standard against which other sources, such as claims data and patient survey information, are compared (Roos 1982, Leatherman 1991). In contrast, physicians' office medical records have escaped such careful scrutiny (Buchsbaum 1987, Moran 1988, Romm 1981, Quam 1993). Similarly, billing information from ambulatory claims has been less frequently studied (Graft 1992, Quam 1993). Because the volume and costs of ambulatory medical care are increasing, understanding the strengths, weaknesses and relationships of its data sources is critical. The purpose of this chapter is to compare information from the physicians' office medical records with information from claims data for a population of Medicare beneficiaries.

Methods

The data for this study come from the DEMPAQ project. The project tested two methods for reviewing office-based care given to Medicare beneficiaries. The project was funded by the Health Care Financing Administration to evaluate techniques that could be used to implement the congressional mandate for office-based review. The project was a joint effort of researchers, physicians and the professional review organizations (PROs) from Maryland, Alabama and Iowa. One review method, using medical record review, was based on medical care functions, such as drug monitoring. The second method used ambulatory and hospital claims to profile practice patterns for chronic diseases, prevention and general office practice. This involved primary care physicians and their patients from Maryland only.

The study sample consisted of Maryland Medicare beneficiaries who had received billed medical services between July 1, 1989, and June 30, 1990. These beneficiaries were assigned to a primary care physician on the basis of the majority of their visits. Primary care physicians included internists, family practitioners and general practitioners as identified by the carriers' provider files. All Maryland physicians who had 25 or more Medicare beneficiaries assigned to them (n=300) were invited to submit copies of a two-year portion (July 1, 1989, through June 30, 1991) of the medical records for a random subset of 20 of their patients. The physician response rate was 30.3% (n = 91) representing 1,998 patients' records.

A parallel set of all claims, including both institutional Part A and B and non institutional Part B claims for these 1,998 patients, were assembled from the Common Working File (CWF) of the National Claims History File (NCH) from the Health Care Financing Administration (HCFA). The raw claims were cleaned and organized, removing repetitive adjudication data and constructing a file that contained a single line for each unique service. Claims data were assembled in January 1992, allowing ample time for all claims to be filed that were related to the review period for this analysis, 12 months from July 1, 1990, through June 30, 1991.

Eight registered nurses from the Delmarva Foundation for Medical Care, Inc., were trained to perform medical record abstraction using a computer-based medical record review system that contained a limited number of items, including demographics, diagnoses, and lab tests and procedures. These items had been identified on the basis of their occurrence in the claims-based practice profiles. A full list of the medical record review items is included in Table 1 (tables are located in Appendix 1). Synonym dictionaries were created and embedded in the computer assisted medical record review format. It is important to note that subjective clinical judgments were not permitted. For example, nurse reviewers were not allowed to assume that elevated blood pressure readings constituted a diagnosis of hypertension if the physician had not specifically indicated that diagnosis. The abstracters were blinded to the occurrence of any of the items in the patients' claims history. After the nurse reviewers' training period, there was an ongoing 5% rereview, with one medical record of each physician assigned to pairs of reviewers.

To compare claims data with medical record data, we created a translation table, specifying the codes from Current Procedural Terminology (CPT-4) for procedures (Coy, 1988) or the International Classification of Diseases (ICD-9 CM) for diagnoses (PMIC) to be considered as matches in the claims files for the items searched in the medical record. These coding definitions were designed to be inclusive. For example, in the medical record we searched for evidence of a blood urea nitrogen (BUN) test. Since such tests are often ordered as part of a test panel, in the claims portion we accepted panel tests as reasonable indicators that the specific test was performed. The translation table is included as Table 2. This decision rule is biased in favor of over identifying matches between claims and medical record information for laboratory tests and procedures.

For matching claims-based diagnoses with those in the primary care physician's medical record, we considered only those diagnoses that occurred on bills submitted by the primary care physician. In contrast, we matched lab tests and procedures regardless of the billing entity, since many such bills are originated by a lab, even though the tests have been ordered by the primary care physician.

Analysis: For yes/no matches, such as with diagnoses, laboratory tests and procedures, we indicate both percent agreement and the kappa statistic (Rosner, 1986). The percent agreement is inflated because of the large number of cases in which the event does not occur in either the medical record or the claims. The kappa coefficient of agreement is the ratio of the proportion of times that the sources agree (corrected for chance agreement) to the maximum proportion of times that the sources could agree (corrected for chance agreement). Its value can range from -1.0 to 1.0. While there is no predetermined level of acceptability, some suggest that a K of less than .40 indicates a poor level of agreement (Fleiss, 1981).

Results

Of the 1,998 medical records, we had *HISKEW* and claims data for 1,927. Analyses for demographics and tests/procedures are based on those 1,927 medical records. Of the 1,927 medical records, only 1,596 had indications of any visit in the 12-month study period, therefore the analysis of diagnoses is based on those 1,596 cases.

Demographics

As shown in Figure 1, the data from Medicare's registration file (HISKEW) and the medical record agree well only for name and gender. Name matched 99.4% of cases when misspellings are considered as matches, and gender matched in 97.7% of cases. For all other demographic data there is a striking lack of agreement. The patient zip code matched in only 40.3% of cases, and date of birth matched in 58.5% of cases. Death status agreed in 34.3% of cases (reviewers were allowed to infer that medical record activity after the end of the study period constituted evidence of survival through the study period). Further inspection of the results in Table 3 shows that most of the lack of matching was due to missing information in the portion of the medical record that was reviewed. This finding suggests that the medical record alone is an insufficient source for providing demographic information. The primary care physician probably maintained a separate registration file that contains some of the missing elements.

Diagnosis

We searched the medical record for the explicit mention of any of 27 different diagnoses any time during the study period. Only 1,596 medical records listed any visits at all during the study period. Matches were not made for specific dates but, rather, considered any mention of the diagnosis in the claims during the study period compared to any mention in the medical record in the same time frame. The percent agreement ranged from 74.9% for hypertension to 99.7% for ketoacidosis (Table 4). However, the percent agreement is inflated since so many cases did not have the disease, thus increasing the match of "not in medical record, nor in claims." As shown in Figure 2, the range for the kappa statistic is from 0.0 to 0.72. The kappa for only six of the diagnoses met or exceeded 0.40. The best agreement, as measured by kappa, was for the most serious major chronic diseases (congestive heart failure, COPD, diabetes, hypertension, rheumatoid arthritis and TIA). Generally, diagnoses appeared more often in the record than in the claims. Notable exceptions to this pattern included potentially severe acute conditions that were more often in the claim than in the medical record, such as acute myocardial infarction, diabetes, hypertension, ischemic heart disease and pneumonia.

The occurrence of these mismatches surprised us, since we had been expecting severe diagnoses to always be noted in the primary care physician's record. Therefore, we conducted a rereview of a small sample of these records. Two reviewers (not from the original eight) independently reviewed 21 medical records. The records were selected because the claims indicated that the beneficiary had a disease while the original medical record review said they did not. (A parallel set of cases where the diagnosis was in the medical record but not in the claims was also analyzed. Those results are not reported here.) Cases were stratified by disease for six diseases of special interest (diabetes, congestive heart failure, COPD, hypertension, ischemic heart disease and osteoarthritis). In this review, the reviewers knew the specific diagnosis that the claims had indicated, and the reviewers were allowed to use supporting clinical evidence to answer the question, "Does this physician think this patient has this disease?" Acceptable evidence included a mention of the specific diagnosis prior to the study time period, or other clinical indicators, such as prescribing micronase and following blood sugar as substantiation of diabetes. When the two reviewers did not agree (two cases), the case was reviewed by a family practitioner.

The results of this rereview are shown in Table 5. In almost half of the cases, the diagnosis had been noted in the previous year, and in an additional three cases there was clinical evidence indicating that the physician was treating the patient for that disease. Data errors were relatively uncommon and went in both directions; that is, in two cases the initial reviewer had missed the diagnosis, and in two cases the claims assignment appeared unwarranted.

Lab Tests and Procedures

In contrast with our uncertainty about the reliability of diagnoses in claims or medical records, we can regard the claims as the gold standard for whether the patient received any lab test or procedure. The strength of this conviction rests on the fact that physicians are paid only for procedures for which they submit bills. In contrast, accurate diagnosis is not required for payment.

As shown in Figure 3 and Table 6, percent agreement ranged from 68.1% to 99.8%, and kappas ranged from -0.02 to 0.73. In general, the correspondence between claims and medical records is worse for tests and procedures than for diagnoses. Not unexpectedly, tests and procedures appeared more frequently in the claims than in the medical record. However, it is more difficult to interpret the "in claims, not in medical record" information because of billing patterns. Claims do not currently carry information about who ordered the test or procedure, only who billed for the test. Therefore, tests could be ordered by some other physician and not found in the primary physician's office medical record. Additional explanations may include that the test was not filed in the medical record, that the report was not copied for review, or even that there was fraud and the test was billed but not done. In any case, these results suggest that there is a lot of testing being conducted that the primary care physician may not know about.

The other type of mismatch in which we have evidence of the test in the medical record but not in the claims is more problematic. Why would a bill not exist for a procedure or lab test that was done? First, the tests could have been done in a hospital and grouped as part of revenue code rather than being billed independently. Second, for tests or procedures that Medicare does not cover (e.g., screening mammograms or influenza immunizations at the time of the study), a bill would not be sent to Medicare. It is also possible that a physician forgot to bill, the patient did not submit the bill to Medicare, or that there were errors in entering the bill into the Medicare system.

Discussion

Caution needs to be exercised in interpreting and generalizing from these results. Most importantly, these analyses are based on records from volunteer physicians who may have better medical records than non volunteers (a separate analysis to compare the characteristics of volunteer and non volunteer physicians is included in Chapter 12). At a minimum, these medical records were more likely to be legible, and the volunteer physicians felt comfortable providing their records for use in developing measures of the quality of health care.

In contrast to the hospital medical record, the ambulatory record cannot yet be considered a gold standard. In the cases where the diagnosis does not appear in the medical record but appears in the claims, the results of this validation study show how infrequently documentation of diagnosis occurs, in spite of the disease being actively managed. It is apparent from our study that common, serious chronic diseases are not frequently explicitly named in physicians' office records. This documentation pattern is distinctly different from that in hospital records and requires a different review strategy. If one is to assess whether the physician is treating the patient for a particular disease, the reviewers must be trained to use consistent clinical judgment in order to draw conclusions about diagnoses. Such a strategy requires a higher level of clinical expertise and a higher level of training to achieve valid and reliable results. These knowledge and training requirements, in turn, make reviewing office records more expensive than is usually anticipated.

Clearly, the claims can indicate much activity the primary physician may not be aware of. Such a finding points to potential new applications of claims for alerting physicians to other care that is being rendered. A common electronic medical record accessible to all providers would provide such information.

The results of this study indicate that claims tend to be highly specific, but not sensitive. If the claim says the person has the disease, the person probably has the disease. However, the claims do not identify all people with the disease. The usefulness of claims depends on the nature of the question being asked. Thus, for many quality assurance purposes claims can be important, since we are often more concerned with rates of certain procedures being done for patients with given conditions (e.g., how many people with diabetes had a HgbA1c), than with how many people had the condition overall. We generally consider it better to give the providers the benefit of the doubt, so we do not want to inflate the denominator with inappropriate cases, resulting in a lower performance rate than is warranted.

> "The more you know what is wrong with a figure, the more useful it becomes."
>
> —*John Tukey*

Appendix 1

Table 1

Review Items for the Medical Record

Demographics

- zip code
- date of birth
- name
- gender
- death

Diagnosis (n = 1596)

- Acute MI
- Angina
- Arthritis, unspecified
- Bronchitis, acute
- Congestive heart failure
- Cardiomegaly
- COPD
- Chest pain
- Diabetes mellitus
- Gangrene
- Gastrointestinal upset
- Hypertension
- Hyperlipedemia
- Hypokalemia
- Ischemic heart disease
- Influenza
- Ketoacidosis
- Neuropathy
- Osteoarthritis
- Pneumonia
- Rheumatoid arthritis
- Diabetic retinopathy
- Renal failure
- Stroke
- Transient ischemic attack
- Peptic or gastric ulcer
- Upper respiratory infection

Lab Test or Procedure (n = 1927)

- Blood urea nitrogen
- Serum calcium
- Colonoscopy
- Serum creatinine
- Chest x-ray
- Debridement
- Electrocardiogram
- Echocardiogram
- Echo stress test
- Ejection fraction
- Exercise tolerance test
- Finger x-ray
- Flu shot
- Foot x-ray
- Blood gases
- Serum glucose
- Hand x-ray
- HGB A1C
- Hip x-ray
- Potassium
- Knee x-ray
- Mammogram
- Sodium
- Nuclear medicine test
- Nucl med card stress test
- Cont oxygen therapy
- Pap smear
- Pulmonary function tests
- Flexible sigmoidoscopy
- Spine x-ray
- Toe x-ray

Table 2

Claims Coding and Medical Records Synonyms for Diagnoses, Procedures and Lab Tests

Diagnosis (n = 1596)	ICD9 Codes for Claims	Synonyms for Medical Record Review
Acute myocardial infarction	410.xx	MI heart attack
Angina	413, 411.1, 413.0	crushing chest pain cardiac chest pain cardiac-type pain
Arthritis	716.xx	unspecified arthritis
Bronchitis	490.xx-491.xx	acute bronchitis DOES NOT INCLUDE CHRONIC BRONCHITIS
Congestive heart failure	402.11, 402.91, 428	heart failure
Cardiomegaly	429.3	enlarged heart
Chronic obstructive pulmonary disease	491.xx, 496.xx, 492.xx	emphysema chronic bronchitis chronic airway obstruction CAO small airways disease chronic obstructive lung disease DOES NOT INCLUDE CHRONIC ASTHMA
Chest pain	786.5	chest pain CP
Diabetes mellitus	250.xx	IDDM NIDDM GDM anything with the word diabetes in it, except "rule-out diabetes" or "r/o diabetes" Type I Type II
Gangrene	785.4	
Gastrointestinal upset	536.8x, 536.9x, 535.xx	gastric distress dyspepsia heartburn nausea vomiting gastric irritation gastritis indigestion

Diagnosis (n = 1596)	ICD9 Codes for Claims	Synonyms for Medical Record Review
Hypertension	401.xx	high blood pressure HBP HASCVD HCVD DOES NOT INCLUDE↑BP
Hyperlipidemia	272.4	hypercholesterolemia elevated cholesterol
Hypokalemia	276.8	low potassium
Ischemic heart disease	414.xx, 411.xx	coronary artery disease coronary atherosclerosis ASCVD
Influenza	487.xx	Flu
Ketoacidosis		diabetic acidosis hyperglycemia ketosis acidosis hyperosmolar, nonketotic coma
Neuropathy	337.xx, 357.xx, 354.xx, 355.xx, 358.1	nerve damage decreased feeling silent autonomic neuropathy diabetic gastroparesis peripheral neuropathy toxic neuropathy mononeuritis carpal tunnel nerve palsy myasthenic syndromes numbness compressive neuropathy paresthesia
Osteoarthritis	715.xx	OA, DJD, degenerative joint disease
Pneumonia	480.xx-486.xx	
Rheumatoid arthritis	714.00-714.29, 714.4, 714.81	RA
Diabetic retinopathy	362.0x, 366.41, 250.5	proliferative retinopathy background retinopathy small vessel disease of retina macular degeneration maculopathy cotton wool spots exudates retinal microaneurysm diabetic cataracts diabetes with ophthalmic manifestations

Diagnosis (n = 1596)	ICD9 Codes for Claims	Synonyms for Medical Record Review
Renal failure	584.xx, 585.xx, 586.xx, 583.81, 581.81, 250.4	kidney failure Kimmelstiel-Wilson's disease acute renal failure chronic renal failure RF diabetic nephropathy diabetic nephrotic syndrome Diabetes with renal manifestations
Stroke	431.xx, 436.xx, 434.xx	CVA cerebrovascular accident cerebral embolism cerebral hemorrhage
Transient ischemic attack	435.9	cerebral insufficiency
Peptic or gastric ulcer	533.xx, 531.xx	ulcer, gastric, or peptic, or unspecified
Upper respiratory infection	465.9x	cold

Lab Test or Procedure (n = 1927)	CPT Codes for Claims	Synonyms for Medical Record Review
Blood urea nitrogen	84520	BUN
Serum Calcium	82310	CA
Colonoscopy	45378, 45380, 45383, 45385	
Serum creatinine	82565	
Chest x-ray	71010-71035	
Debridement		cleaning wound cleaning
Electrocardiogram		EKG tracing
Echocardiogram	93307-93325	echocardiography doppler echocardiography
Echocardiography stress test	93350	echocardiography with stress test echo with bicycle exercise
Ejection fraction	78414	
Exercise stress test	93015-93018	exercise tolerance test
Finger x-ray	73140	
Flu shot	90724	influenza immunization
Foot x-ray	73620, 73630	
Blood gases	82790-82817, 94700, 94705, 94710	
Serum glucose	82948, 82951	blood sugar, BS
Hand x-ray	73120-73130	

Lab Test or Procedure (n = 1927)	CPT Codes for Claims	Synonyms for Medical Record Review
HGB A1C	83036	hgbA1c glycosylated hemoglobin (hb or hgb) A1c hemoglobin A1 glycated hemoglobin
Hip x-ray	73500-73531	
Potassium	84132	
Knee x-ray	73560-73581	
Mammogram	76090-76091	
Sodium	84295	
Nuclear medicine test	78415, 78460, 78466-78469, 78470-78474, 78481-78484	radionuclide scan gated cardiac scan
Nuclear medicine cardiac stress test	78461, 78476-78477, 78486-78487	radionuclide exercise tolerance test exercise thallium test gated cardiac imaging with exercise
Continuous oxygen therapy	CSF 87c - codes A, I, P, or R; E1400, E0430, E0425, Q0043, Q0046, E0410, E0416	continuous oxygen therapy continuous O2
Pap smear	88150, 88151, 88155	
Pulmonary function test	94010, 94060, 94070	spirometry
Flexible sigmoidoscopy	45330-45333, 45336, 45337	
Spine x-ray	72010-72120	
Toe x-ray	73660	

Figure 1

Correspondence of Demographics Between Medical Record and HISKEW File

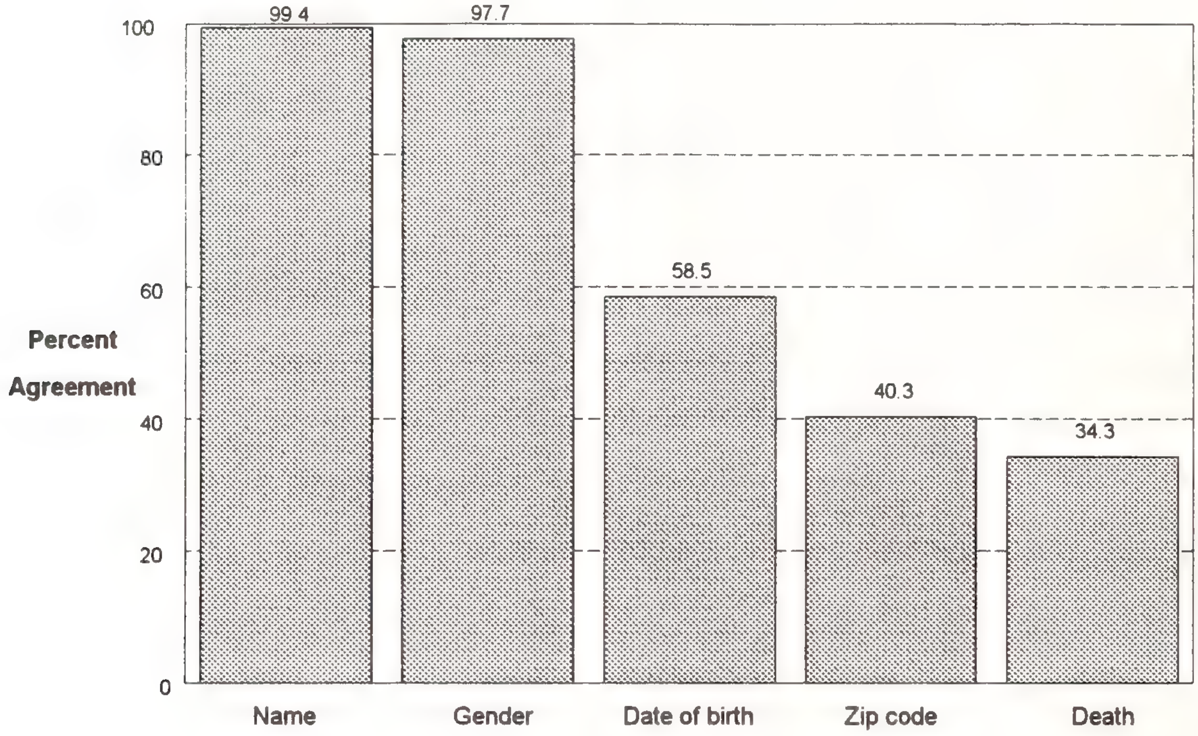
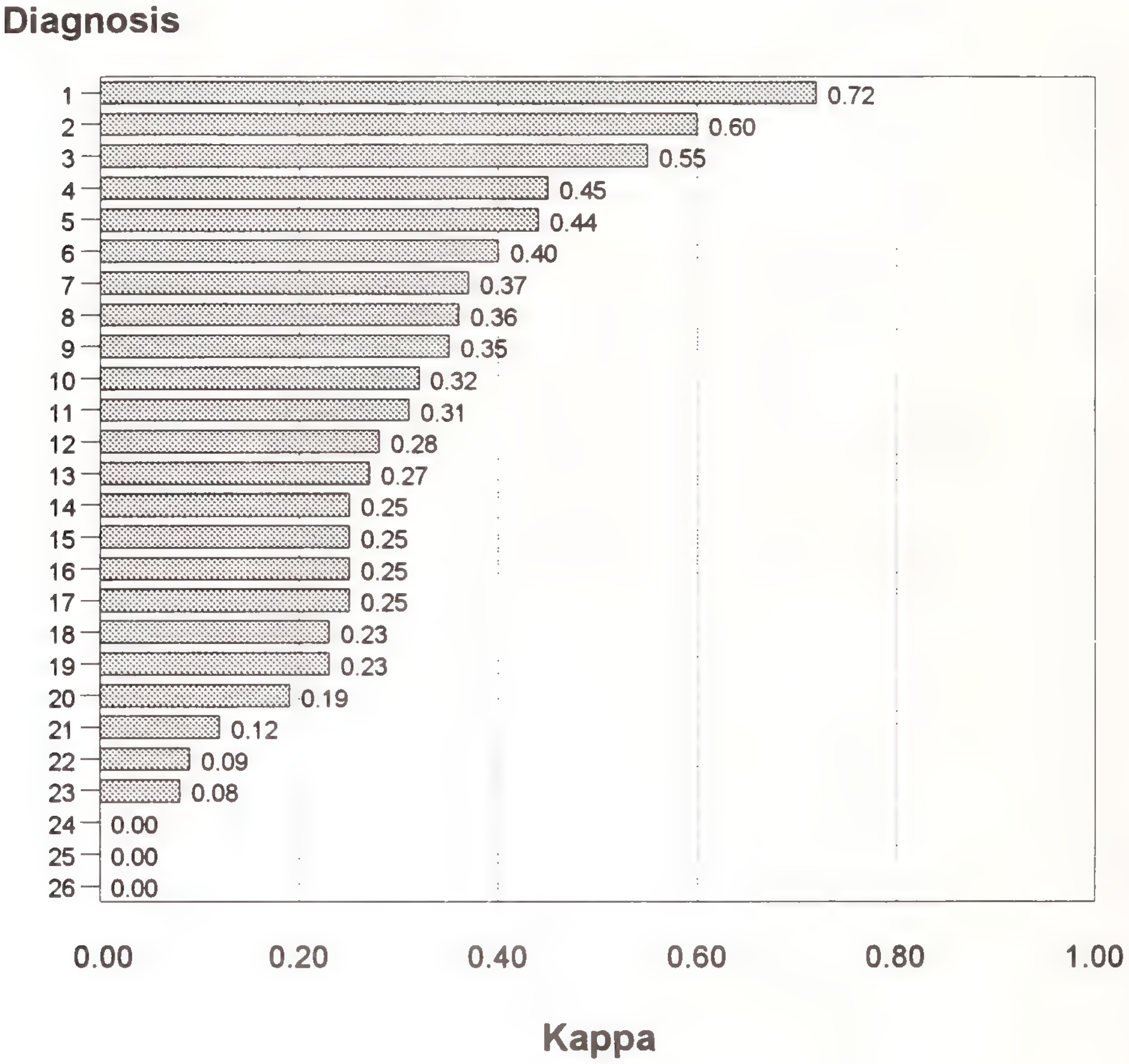


Table 3
Demographics
(n = 1927)

Zip code:			
	Match	776	40.3%
	No match:	1151	59.7%
	In medical record, no match	87	4.5%
	Missing from medical record	1064	55.2%
Date of birth:			
	Match	1127	58.5%
	No match:	800	41.5%
	In medical record, no match	210	10.9%
	(of these 210, >364 days	94	4.9%
	Missing from medical record	590	30.6%
Death:			
	Match	660	34.3%
	No Match	1267	65.7%
	not coded in medical record review	2	0.1%
	don't know in medical record	1260	65.4%
	in medical record, not in HISKEW	3	0.2%
	in HISKEW, not in medical record	2	0.2%
Name:			
	Match	1854	96.2%
	No match	73	3.8%
	intelligent match	62	
	missing in HISKEW	8	
	egregious	3	
Gender:			
	Match	1883	97.7%
	No match	44	2.3%
	"2" in HISKEW, "1" in medical record	19	
	"1" in HISKEW, "2" in medical record	12	
	missing in medical record	13	

Figure 2

Correspondence Between Diagnosis in Primary Physician's Medical Records and Claims



Diagnoses:

1= Diabetes mellitus
2= Congestive heart failure
3= COPD
4= Hypertension
5= Rheumatoid arthritis
6= Transient ischemic attack
7= Angina
8= Pneumonia
9= Acute MI
10= Upper respiratory infection
11= Osteoarthritis
12= Peptic or gastric ulcer
13= Ischemic heart disease
14= Arthritis, unspecified
15= Chest pain
16= Hyperlipidemia
17= Stroke
18= Bronchitis, acute
19= Influenza
20= Gastrointestinal upset
21= Cardiomegaly
22= Renal failure
23= Neuropathy
24= Hypokalemia
25= Ketoacidosis
26= Diabetic retinopathy

Table 4

Agreement Between Diagnosis in Primary Physician's Medical Records[1] and Claims[2]

Diagnosis (n = 1596)	Kappa	Agreement: In Chart and Claims	Agreement: Not in Chart Nor in Claims	Disagreement: In Claims Not in Chart	Disagreement: In Chart Not in Claims	Percent Agreement
Diabetes mellitus	0.72	175	1313	54	54	93.2
Congestive heart failure	0.60	58	1468	29	41	95.6
COPD	0.55	74	1416	37	69	93.4
Hypertension	0.45	361	835	203	197	74.9
Rheumatoid arthritis	0.44	4	1582	5	5	99.4
Transient ischemic attack	0.40	5	1576	4	11	99.1
Angina	0.37	23	1500	18	55	95.4
Pneumonia	0.36	6	1569	14	7	98.7
Acute MI	0.35	3	1582	8	3	99.3
Upper respiratory infection	0.32	52	1365	19	160	88.8
Osteoarthritis	0.31	59	1343	71	123	87.8
Peptic or gastric ulcer	0.28	6	1560	9	21	98.1
Ischemic heart disease	0.27	55	1331	100	110	86.8
Arthritis, unspecified	0.25	46	1341	31	178	86.9
Chest pain	0.25	21	1467	32	76	93.2
Hyperlipidemia	0.25	32	1411	37	116	90.4
Stroke	0.25	5	1562	5	24	98.2
Bronchitis, acute	0.23	23	1447	25	101	92.1
Influenza	0.23	4	1566	16	10	98.4
Gastrointestinal upset	0.19	16	1464	11	105	92.7
Cardiomegaly	0.12	1	1581	0	14	99.1
Renal failure	0.09	1	1577	7	11	98.9
Neuropathy	0.08	3	1533	4	56	96.2
Hypokalemia	0.00	0	1585	4	7	99.3
Ketoacidosis	0.00	0	1591	4	1	99.7
Diabetic retinopathy	0.00	0	1584	2	10	99.2
Gangrene	(na)	0	1596	0	0	100.0

[1] Diagnosis recorded at a visit at least once for the 12 months 7/1/90-6/30/91
[2] Face-to-face visits billed by that physician

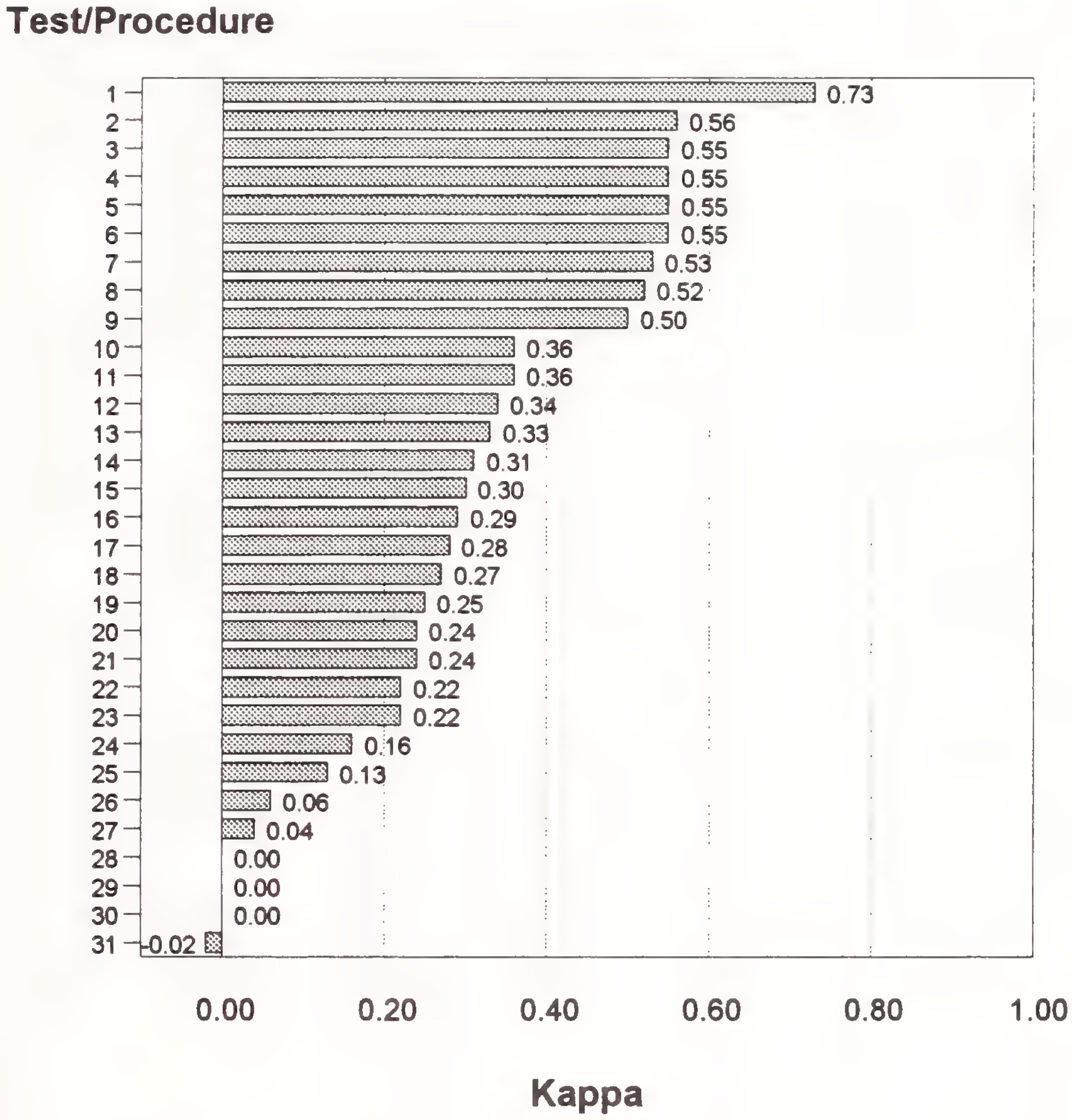
Table 5

Results from Re-Review of Diagnoses

Number of Cases (n=21)	Percent	Category
10	47.6%	Diagnosis in previous year
3	14.3%	Clinical evidence
2	9.5%	More general diagnosis
2	9.5%	Illegible
2	9.5%	MD not treating disease
1	4.8%	On Problem List
1	4.8%	Missed on initial review

Figure 3

Correspondence Between Lab Tests & Procedures in Primary Physician's Medical Records and Claims



Lab Tests/Procedures:

1= HGB A1C
2= Serum creatinine
3= Blood urea nitrogen
4= Serum calcium
5= Potassium
6= Sodium
7= Mammogram
8= Serum glucose
9= Cont oxygen therapy
10= Flu shot
11= Spine x-ray
12= Electrocardiogram
13= Exercise tolerance test
14= Pulmonary function test
15= Nuc med card stress test
16= Pap smear
17= Flexible sigmoidoscopy
18= Colonoscopy
19= Chest x-ray
20= Hip x-ray
21= Knee x-ray
22= Finger x-ray
23= Foot x-ray
24= Echocardiogram
25= Blood gases
26= Hand x-ray
27= Debridement
28= Echo stress test
29= Ejection fraction
30= Toe x-ray
31= Nuclear medicine test

Table 6

Agreement Between Lab Tests & Procedures in Primary Physician's Medical Records and Claims

		Agreement		Disagreement		
Lab test or Procedure (n = 1927)	Kappa	In Chart And Claims	Not in Chart Nor in Claims	In Claims Not in Chart	In Chart Not in Claims	Percent Agreement
HGB A1C	0.73	54	1835	24	14	98.0
Serum creatinine	0.56	608	902	266	151	78.4
Blood urea nitrogen	0.55	616	888	268	155	78.0
Serum calcium	0.55	572	936	293	126	78.3
Potassium	0.55	628	871	273	155	77.8
Sodium	0.55	600	903	272	152	78.0
Mammogram	0.53	126	1614	171	16	90.3
Serum glucose	0.52	674	784	317	152	75.7
Cont oxygen therapy	0.50	9	1900	17	1	99.1
Flu shot	0.36	113	1528	87	199	85.2
Spine x-ray	0.36	41	1753	124	9	93.1
Electrocardiogram	0.34	325	988	584	30	68.1
Exercise tolerance test	0.33	21	1826	74	6	95.8
Pulmonary function tests	0.31	17	1837	70	3	96.2
Nucl med card stress test	0.30	9	1877	37	4	97.9
Pap smear	0.29	29	1773	90	35	93.5
Flexible sigmoidoscopy	0.28	11	1861	49	6	97.1
Colonoscopy	0.27	15	1835	76	1	96.0
Chest x-ray	0.25	171	1155	589	12	68.8
Hip x-ray	0.24	13	1836	77	1	96.0
Knee x-ray	0.24	14	1828	82	3	95.6
Finger x-ray	0.22	1	1919	7	0	99.6
Foot x-ray	0.22	10	1851	65	1	96.6
Echocardiogram	0.16	18	1741	167	1	91.3
Blood gases	0.13	4	1870	50	3	97.2
Hand x-ray	0.06	1	1894	31	1	98.3
Debridement	0.04	1	1880	41	5	97.6
Echo stress test	0.00	0	1918	6	3	99.5
Ejection fraction	0.00	0	1924	0	3	99.8
Toe x-ray	0.00	0	1923	4	0	99.8
Nuclear medicine test	-0.02	0	1857	30	40	96.4

Chapter 10
Evaluating the DCRS: Assessment by Physicians

Abstract
We examined several aspects of the validity of the claims profiles, i.e., how well do the profiles measure quality? We conducted two evaluations of the face validity of the profiles and one assessment of the validity of the profile components. The latter study compares the interpretation of the profiling results for each profile component against an interpretation derived from the office record.

First Assessment of Face Validity
The first test of face validity took place during the development process with the dissemination of draft profiles to liaison physicians and data experts. This initial review led to changes in the profile specifications.

Second Assessment of Face Validity
A second, more structured, test of face validity was undertaken during the profile component evaluation and was limited to the profiles of diabetes, CHF, COPD and preventive services. Seven Maryland physician reviewers completed structured worksheets which asked questions about the physicians' perceptions of the validity of selected profile components. This review confirmed the face validity of the component pieces of the selected profiles.

Structured Profile Component Evaluation
The evaluation of the profile components was a structured review where the physician reviewers were asked about the appropriateness of the care received by the patient. In essence, this study compared the interpretation of data based on the claims with the interpretation of data based on the office record. This review suggests that the presence of a claim for a recommended care item indicates good quality care. If the claim is not found, three categories of reasons appear to explain its absence: the patient didn't need the service; the service may have been provided outside the Medicare system, e.g., in the VA system; and the patient may not have received a needed service. The first explanation is particularly helpful for understanding and interpreting the limited use data.

Discussion
To maintain credibility of performance measures, it is important to undertake periodic validation of the measures. Medical practice changes rapidly and items which had face validity a few years ago, may not have validity today. Both unstructured and structured tests prove useful for assessing face validity. A structured peer review of profile components also measures validity, producing results that can be used to clarify and modify the interpretations of the profiles and can lead to improvements in the performance measures.

This chapter presents the results of several tests of the validity of the claims profiles. We first give an example of why we need to test validity. Next, we describe the questions to be addressed by the evaluation. Third, we present the methods and results for each of the validity studies, and we conclude with a discussion of the implications of our research.

The main goal of this analysis was to see if the message the claims data tells us is correct. Assume, for example, we do not find a claim for an eye exam for a diabetic patient in the study year. We might infer from the claims that this patient does not receive good quality of care. However, a physician looking at the office record may find any number of reasonable explanations for the exam's absence. The patient may be blind, or may be terminally ill. The primary care physician may have dilated the pupil in his/her office and not billed separately for the procedure. It is crucial to know how often these alternative explanations occur. Moreover, are ophthalmic exams even an important measure to consider? The rest of this chapter presents a more formal description of how we studied these issues.

Research Questions

Validity refers to how well a tool measures the concept it is supposed to measure. In the case of the DCRS profiles, the validity question is how well do the profiles measure quality? Validity may be appraised several ways. We consider only two here. ***Face validity*** is whether the measures *appear* to a knowledgeable person to make sense (Abramson, 1990). ***Criterion validity*** is how well the results of the measurement compare to a gold standard (Abramson, 1990). While a gold standard for quality does not exist, the concept of comparing the interpretation of the profiling results against an interpretation derived from another data source proves useful. This comparison allowed us to identify various factors which affect the interpretation and validity of the profiles.

We designed the physician assessment of profiles to answer the following questions:

- Do the profiles include services which physicians agree are important for the care of patients with specific conditions, i.e., do the profiles have face validity?
- How valid are inferences about the quality of medical care received by a group of patients, based on performance rates derived from individual services? What patient, practice or other factors that could be discovered in the office record explain why a particular service does or does not appear in the claims database and therefore affect the interpretation of the claims data?

First Assessment of Face Validity:
Liaison Physicians (All States) and Data Experts (National)

The face validity of the profiles was considered twice during the course of the project. The first assessment of face validity took place during the development phase of the project when the initial profile specifications were shared with liaison physicians and data experts. The second evaluation of face validity took place in a more structured review setting where physicians were asked to complete a questionnaire with specific questions about the validity of the measures. This section describes the first evaluation.

Methods: First Assessment of Face Validity

Profile specifications were presented to liaison physicians during special briefing sessions designed to introduce the claims profiling. Liaison physicians were individuals in each of the three study states who had been identified by their specialty societies to serve as a link between the physician community and the project. The liaison meetings took place in the fall and winter of 1991-1992 in each of the three study states. At these briefing sessions the theoretical underpinnings of the profiling were presented and a variety of technical issues was discussed. Copies of the profile specifications were distributed to the physicians for their critical review and comment.

Profile specifications were also mailed to selected nationally recognized data experts for their review and comment. The experts were invited to comment based on their expertise with large claims data sets, with case-mix or severity adjustment, or with Medicare data.

Table 1, on the following page, shows an example of the condition-specific and prevention profile specification format. The first column listed a service that was proposed for inclusion in the profile. The next column listed the CPT or ICD-9 codes that defined the service. The final column provided a description of the CPT or ICD-9 codes.

Table One

Example from Diabetes Profile specification worksheet

Profile Component	CPT or ICD9 codes	Description
Ophthalmology Visit, annual	92002	Ophthalmological services: medical examination and evaluation with initiation of diagnostic and treatment program; intermediate, new patient
	92004	Ophthalmological services: medical examination and evaluation with initiation of diagnostic and treatment program; comprehensive, new patient, one or more visits
	92012	Ophthalmological services: medical examination and evaluation with initiation of diagnostic and treatment program; intermediate, established patient
	92014	Ophthalmological services: medical examination and evaluation with initiation of diagnostic and treatment program; comprehensive, established patient, one or more visits
	92019	Ophthalmological examination and evaluation, under general anesthesia, with or without manipulation of globe for passive range of motion or other manipulation to facilitate diagnostic examination; limited
	92225	Ophthalmoscopy, extended as for retinal detachment (may include use of contact lens, drawing or sketch, and/or fundus biomicroscopy), with medical diagnostic evaluation; initial
	92226	Ophthalmoscopy, extended as for retinal detachment (may include use of contact lens, drawing or sketch, and/or fundus biomicroscopy), with medical diagnostic evaluation; subsequent
	92230	Ophthalmoscopy, with medical diagnostic evaluation; with fluorescein angioscopy (observation only)
	92235	Ophthalmoscopy, with medical diagnostic evaluation; with fluorescein angiography (includes multiframe photography)
	92250	Ophthalmoscopy, with medical diagnostic evaluation; with fundus photography
	92260	Ophthalmoscopy, with medical diagnostic evaluation; with ophthalmodynamometry
OR office visit with ophthalmologist	90000-90080 and 90600-90654	Office visit, if specialty of physician is ophthalmology

Liaison physicians and data experts were asked to comment on the importance of each service for the care of an individual with the specified condition, or for the prevention of morbidity in the case of the prevention profile, and on the appropriateness of the coding.

The specification format for the office practice profile consisted of a series of draft tables. Text accompanying the specification explained that the final report to physicians would include narrative on the facing pages to aid in interpreting the tables. Comments were solicited on the general interest and usefulness of this type of data for practicing physicians.

Results: First Assessment of Face Validity

The complete summary of liaison and data expert comments is shown in Appendix 1. Most of the issues raised by the liaison physicians and data experts were addressed prior to running the final profiles. Many services were eliminated or revised based on the feedback received at this stage of the project. Helpful suggestions included guidance about severity stratification, codes used to define the services, and table format.

Second Assessment: Structured Review of Face Validity—Maryland Only

The Delmarva Foundation for Medical Care, Inc., undertook an additional test of face validity as part of the "profile component evaluation study" described later in this chapter. This test was limited to selected condition-specific profiles: diabetes; congestive heart failure; chronic obstructive pulmonary disease; and prevention. As profile development was nearly complete at the time of this review, physicians were asked to consider not only whether the services should be included in the profile, but whether the clinical rationale for including the services was appropriate and whether a stratification of patients based on severity was required.

Seven physicians recruited from the Maryland primary care specialty societies were given a set of tables, each table addressing one of the services listed in the profile (see Table 2 on the next page). Physician reviewers were asked to consider the following questions:

- Is the inclusion of this component appropriate for this profile? Should it be retained, eliminated, or revised?
- Is the classification of the component as "recommended" or "limited use" appropriate?
- Is the rationale correct? Should it be revised?
- Is the severity stratification correct? Should it be revised?

A portion of the table worksheet is shown on the next page; full copies summarizing the physicians' comments are provided in Appendix 2. There is a column on the table for each of the previous questions. The top row shows the proposed service (called a "component"), its attribute, rationale and severity stratification. If physician reviewers disagreed with the choice or use of the component, they were asked to explain their views in the space provided.

Table Two

Structured Face Validity Worksheet

Congestive Heart Failure

Recommended Care				
Office Visit	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a minimum of two office visits within the study year, one in each half year.	Recommended	The primary care visit for patients with CHF is important in monitoring drug does, drug side effects, or symptoms of heart failure. Most primary care practitioners schedule office visits 2 - 4 months apart (Fleg et al, 1989). For this profile, 2 visits per year are considered minimal care.	None
Suggested Revisions				

Physician reviewers were also asked to suggest other services that should be included in the profile. They were reminded that many important quality indicators such as medications or patient education efforts do not appear in the claims data and thus could not be included in the claims profiles.

Results: Second Assessment of Face Validity
Appendix 2 presents a complete summary of the results of the structured face validity study. Overall, the physicians confirmed the validity of including the chosen services in the condition-specific profiles and offered specific suggestions for improving the profiles. Many of the changes they suggested were incorporated into the final version of the profile specifications. However, the physicians frequently forgot that certain types of information are unobtainable from the claims, e.g., shortness of breath, results of ECG tracings, or results of cholesterol measurements. The following summarizes the key recommendations of the physicians.

Diabetes. One reviewer questioned the validity of designating ophthalmology visits, triglycerides, HDL cholesterol and total cholesterol as recommended care services. Since the suggestion regarding ophthalmology visits is counter to recent literature stressing the importance of annual eye exams for diabetic patients, we left ophthalmology visits under recommended care (Brechner, 1993). The lipid studies were also left under recommended care because of the diabetic patient's risk of heart disease.

Congestive Heart Failure. The reviewers confirmed the validity of including the various services in the CHF profile. Many small changes were proposed for the clinical rationales. Substantive changes were proposed for two services, echocardiography and ECG. For echocardiography, reviewers proposed adding a severity stratification: patients with valve disease, ventricular dysfunction, pericardial effusion, cardiac mass. While this change was feasible (ICD-9 codes are present for these conditions), technically it proved extremely difficult to stratify the data set for a single service. For ECG, two reviewers suggested making this a recommended care item. We left it as a limited use item because, as a third reviewer pointed out, ECGs are not necessary if the patient has stable CHF.

Chronic Obstructive Pulmonary Disease. The reviewers questioned the validity of including several of the services in the COPD profile, noting that "not all patients require these services." The services questioned by the physician reviewers were retained in the profile as limited use services.

Preventive Services. For several of the services, the reviewers suggested changing the attribute from limited use to recommended. These services were total cholesterol, pap smear, sigmoidoscopy and PSA. These changes were not made because the preventive services profile was later dropped.

Structured Profile Component Evaluation

A study of the validity of the inferences drawn from the profiles was undertaken in Maryland.[1] The inferences concerned the quality of primary care rendered to Medicare beneficiaries. This study compared the interpretation of the data based on the claims to an interpretation of the data based on the office record. For example, if the claim suggested that a service was not provided to the patient, did that imply that the patient received poor quality care, or were there other factors that could be discovered in the office record to explain why the service was not provided?

Selecting profiles for review. This review focused on three of the condition-specific profiles—congestive heart failure, diabetes, and chronic obstructive pulmonary disease—and the preventive services profile. The office practice profile was excluded from the evaluation because it lacked clinical information which could be peer reviewed by physicians. The number of profiles to review was further narrowed by excluding profiles for which the clinical data were limited, such as hypertension and osteoarthritis. The final decision of which profiles to review was made on the basis of visit volume.

Selecting components for review. We limited our analysis to the recommended care and limited use items because:

- Not all services could be found in the office record. For example, we expected recommended care items to be present in the record; limited use items to be present only some of the time. Similarly, only some of a patient's hospitalizations may be found in the office record.
- We had limited resources for conducting the review.

Sampling for review. For the three condition-specific profiles, the sampling frame consisted of patients for whom medical records had been previously submitted (see Chapter 11), with the conditions of diabetes, congestive heart failure and chronic obstructive pulmonary disease. The conditions were determined by searching the claims for patients who had had face-to-face visits with their primary care source where the principal diagnosis was the condition of interest. This was to increase the likelihood that medical records would contain information about the selected condition. Fifty patients with each condition were randomly selected.

The preventive services sampling frame consisted of any patient for whom medical records had been submitted, where the patient had had a face-to-face visit with their primary care source during the study year, July 1, 1990, through June 30, 1991. Fifty patients were randomly selected for the preventive services profile.

Profile component evaluation form. We developed a single page form with a self-adhesive, computer-generated label identifying the profile and component to be reviewed. For example, the label for diabetes eye care listed the profile as "Diabetes" and the component as "Ophthalmology Visit." There were four possible worksheets, depending on whether the component was ***Recommended*** or ***Limited Use***, and whether the care was ***done*** or ***not done.*** Each worksheet asked a different set of questions to determine whether a profile component was useful as a quality indicator *for that case only.* The four worksheets are shown in Appendix 3. The first part of the worksheet asked the physician to confirm that the patient had the diagnosis. The second part asked the following types of questions.

For recommended care:

- ***If the care was done,*** are there any factors in the medical record that suggest this care should not have been done for this patient?

---

[1] Funding and time constraints limited the test to a single site.

- ***If the care was not done,*** are there any reasons given in the medical record which explain why the care could or should not be given for this case? Such reasons include "Patient Reasons": refusal, terminal illness; "Practice Reasons": instrumentation not available; or "Extenuating Circumstances": a blizzard prevents the patient from reaching the physician.

For limited use care:

- ***If the care was done,*** do the facts in the office record justify the care given? If justified, what diagnoses or procedures might be used as evidence for the justification? If not justified, what patient, practice or other factors explain why the care should not have been given?
- ***If the care was not done,*** do the facts of the case justify the absence of care? What are the patient, practice or other factors which explain why the care should not be done?

Conducting the review. First, the nurse reviewer selected a worksheet, based on the information on the label and affixed the label to the worksheet. Next the nurse examined the chart to become familiar with its contents. During the physician review sessions which followed, the nurse was available to answer questions about the purpose of the evaluation, to assist the physician advisor in finding data in the chart, and to ensure that the review form was correctly filled out.

Following the completion of the review, the data were entered into a Paradox database.

Results: Structured Profile Component Evaluation

Cases reviewed. Initially, 50 office records were examined for each profile. Several months after the completion of data collection, we realized that the programming code for selecting a patient with a condition had been faulty. We went back to the data and excluded those patients or beneficiaries who did not have a face-to-face visit with their primary care source for the condition of interest. This reduced the samples to:

Diabetes	30
Congestive Heart Failure	29
Chronic Obstructive Pulmonary Disease	28

The preventive services profile sample remained at 50.

Findings. In general, the profile component evaluation confirmed the validity of the recommended care items and aided in the interpretation of the limited use items. The sample of cases is so small, however, that the results presented here must be viewed as preliminary rather as definitive findings.

Recommended care items. The presence of a claim always indicated good quality care for the recommended care items, i.e., the physician reviewers confirmed the appropriateness of the care (see Table 3 on the next page).

Table Three
Profile Component Evaluation
Recommended Care Items
Summary Results, Across All Profiles

Component	Cases where a claim indicates good quality		Cases where "no claim" indicates			
				good quality	care not billed to Medicare	POSSIBLE poor quality care
	N	%	N	%	%	%
DM, CHF, COPD: Office Visit	85	100%	2	0%	100%	0%
DM, CHF, COPD, Prev: Influenza Immunization	21	100%	82	7%	22%	71%
DM: HgbA1c	7	100%	23	0%	0%	100%
DM: HDL Cholesterol	5	100%	17*	18%	0%	82%
DM: Total Cholesterol	7	100%	13*	31%	0%	69%
DM: Urinalysis	18	100%	12	0%	17%	83%
DM: Ophthalmology Visit	12	100%	18	11%	6%	83%
DM: Triglycerides	8	100%	11*	27%	0%	73%
Prevention: Mammogram	0	NA	18	11%	11%	78%

*** Coding problem cases were excluded from denominator of this table**

Sometimes, the medical record showed that the patient had, indeed, received the service, although no claim had been generated. This occurred when the patient either received care outside the Medicare system, such as through the VA or public health system, or where the charge for the service was included in a global fee, such as for a hospitalization or office visit. Table 4 on the next page lists the services where the office record revealed that the patient actually did receive care, but it was not billed separately through Medicare.

Table Four
Recommended Services not billed to Medicare

Profile: Service	N No Claim	Percent of cases *without a claim*, where patient received care, but care not billed to Medicare	Where patient received services, or payer
Diabetes: Office Visit	1	100%	Medicaid
Diabetes: Flu Shot	29	42%	Public Health Department
Diabetes: Urinalysis	12	17%	Included in office charge
Diabetes: Ophthalmology visit	18	6%	Medicaid
CHF: Office Visit	1	100%	VA
CHF: Flu Shot	27	11%	VA, Public Health Department
COPD: Flu Shot	20	15%	Public Health Department
Prevention: Mammogram	18	11%	No mention where
Prevention: Flu Shot	29	15%	Probably included in office charge, or from Public Health Department

Limited use items. The presence of a claim, usually, although not always, indicated good quality care for limited use care items. Occasionally the physician reviewer thought the patient had received a service they did not need (see Appendix 4). The problem services were CXR for CHF or COPD (two cases) and Home O_2 for COPD (one case).

The absence of a claim generated three explanations: the service was received outside the Medicare system, the patient did not need the service, or the patient may have required the service but did not get it. Table 5 on the next page shows limited use services which are not billed separately or directly to Medicare.

Table Five
Limited Use Services not billed to Medicare

Profile: Service	N No Claim	Percent of cases *without a claim,* where patient received care, but care not billed to Medicare	Where patient received services, or payer
CHF: ECG, Chest X-ray	10	50%	VA, Hospital
CHF: Echocardiography	17	18%	VA, Hospital
CHF: Stress Test	27	4%	Hospital
COPD: Blood gases	21	30%	Hospital

For many limited use services, the absence of a claim also meant the patient was receiving good quality care. For example, in 86% of the cases where no claim was filed for angiography, the patient's clinical condition did not warrant the procedure. Table 6 shows selected limited use items where no claim indicates the patients with the condition did not require the service.

Table Six
Selected Limited Use Items
Cases where NO claim = Good Quality

Profile: Limited Use Item	N No Claim	Percent of cases where physician reviewers found that no claim indicates good quality care
CHF: Angiography	28	86%
CHF: Echocardioghraphy	17	47%
CHF: Stress Test	27	67%
COPD: Blood gases	21	42%
COPD: Chest Xray	6	83%
COPD: Spirometry	17	47%
COPD: Home O2	23	73%
Prevention: Colonoscopy	48	40%
Prevention: PSA	17	41%

The remainder of the cases without a claim may include situations where the patient did not receive appropriate care. Lack of complete medical records from all providers of health care to the patient made it impossible to definitively determine whether the patient did not receive needed care.

Table Seven
Profile Component Evaluation: DIABETES
Summary Results

Component	Cases where a claim indicates good quality		Cases where "no claim" indicates			
				good quality	care not billed to Medicare	POSSIBLE poor quality
	N	%	N	%	%	%
Office Visit	29	100%	1	0%	100% (Medicaid)	0%
Influenza Immunization	1	100%	29	10%	42% (Public Health)	48%
HgbA1c	7	100%	23	0%	0%	100%
HDL Cholesterol	5	100%	17*	18%	0%	82%
Total Cholesterol	7	100%	13*	31%	0%	69%
Urinalysis	18	100%	12	0%	17% (Probably included in office charge)	83%
Ophthalmology Visit	12	100%	18	11%	6% (Medicaid)	83%
Triglycerides	8	100%	11*	27%	0%	73%
Blood Glucose	22	100%	3*	33%	0%	67%

*** Coding problem cases were excluded from denominator of this table**

Table 7 above and Tables 8, 9, and 10 on the following pages show the complete summary results of the profile component evaluation. Additional data are provided in Appendix 4.

Table Eight
Profile Component Evaluation: CHF
Summary Results

Component	Cases where a claim indicates good quality		Cases where "no claim" indicates			
				good quality	care not billed to Medicare	POSSIBLE poor quality
	N	%	N	%	%	%
Office Visit	28	100%	1	0%	100% (VA)	0%
Influenza Immunization	2	100%	27	4%	11% (VA, Public Health)	85%
Potassium	2	100%	9*	22%	0%	78%
Angiography	1	100%	28	86%	0%	14%
ECG	23	87%	6	33%	50% (Hospital, VA)	17%
Chest X-ray	25	88%	4	25%	50% (Hospital, VA)	25%
Echocardiography	12	83%	17	47%	18% (Hospital, VA)	35%
Stress Test	2	100%	27	67%	4% (Hospital)	29%

* Coding problem cases were excluded from the denominator of this table

Table Nine
Profile Component Evaluation: COPD
Summary Results

Component	Cases where a claim indicates good quality		Cases where "no claim" indicates			
				good quality	care not billed to Medicare	POSSIBLE poor quality
	N	%	N	%	%	%
Office Visit	28	100%	0	0%	0%	100%
Influenza Immunization	20	100%	8	5%	15% (Public Health)	80%
Blood gases	21	86%	7	42%	38% (Hospital)	19%
Chest x-ray	22	91%	6	83%	0%	17%
Spirometry	11	100%	17	47%	0%	52%
Home O2	5	80%	23	73%	0%	26%

Table Ten
Profile Component Evaluation: Prevention
Summary Results

Component	Cases where a claim indicates good quality		Cases where "no claim" indicates			
				good quality	care not billed to Medicare	POSSIBLE poor quality
	N	%	N	%	%	%
Influenza Immunization	10	100%	40	8%	15% (Public Health, office charge)	78%
Mammogram	0	NA	18	11%	11% (?? where)	78%
Sigmoidoscopy	2	100%	48	38%	0%	62%
Pap Smear	2	100%	31	23%	0%	77%
Colonoscopy	1	100%	48	40%	0%	60%
PSA	0	NA	17	41%	0%	59%
Total Cholesterol	5	100%	45	33%	0%	67%

Using the Results to Interpret Data

These data may provide many useful insights for the interpretation of the profile data. We will use examples from several profiles to illustrate how the data may be used to interpret the recommended and limited use results. The number of cases in our study are quite small and therefore the examples presented here should be viewed as hypothetical.

Recommended Care. If a claim is found for a recommended care item, the physician reviewers concluded that the care was necessary and appropriate. For example, all flu shots received by the sampled patients were deemed appropriate. However, the converse is not true. A missing flu shot claim does not necessarily mean that the patient did not receive appropriate care. Let's assume that 80% of all diabetic patients did not receive a flu shot. Based on the information presented above in Table 7, we discover that for 10% of the 80% (8% of all diabetics), a flu shot is either refused or not given because of patient allergies. Forty-two percent of those without a flu shot (34% of diabetics: .42x.8) received their flu shot elsewhere, through the Department of Public Health, the VA, etc. This leaves 38% of all diabetics (.48x.8) not receiving a flu shot who perhaps should have received one.

Limited Use Care. Tables 7-10 are perhaps most helpful in interpreting limited use items. Limited use care, by definition, applies to some (a limited number) patients with the condition. The difficulty is determining, through the claims, which patients need the service.

If a claim is found for a limited use item, the care was usually deemed necessary and appropriate by the physician reviewers. For example, 83% of all echocardiographies received by CHF patients were deemed appropriate. The other 17% lacked enough documentation in the record to determine whether the provision of the service was necessary or not. Thus, if 45% of all CHF patients receive an echocardiogram, then perhaps as many as 8% did not require the service (.17x.45).

The absence of a claim for a limited use service is often explained by patient clinical factors. Using the echocardiography example described above, Table 8 indicates that 47% of the CHF patients for whom no claim was located (21% of all CHF patients—.47x.45), lacked a claim because the patient's clinical condition did not warrant the provision of the service. Eight percent of the CHF patients (.18x.45) without a service received care at the VA or through a hospital. Although Medicare was billed for the hospital service, it was included in a lump sum charge and not broken out by test. VA services are outside the Medicare system. This leaves 16% of all CHF patients (.35x.45) who may have needed an echocardiogram and did not receive one.

Summary and Recommendations

To maintain credibility of performance measures, it is important to undertake periodic validation of the measures. Medical practice changes rapidly and items which had face validity a few years ago may not have validity today. Both unstructured and structured tests prove useful for assessing face validity. A structured peer review of profile components also measures validity, producing results that can be used to clarify and modify the interpretations of the profiles and can lead to improvements in the performance measures.

We recommend that:

1. Face validity checks be undertaken periodically to assess the value of including particular services in the profile.

2. The profile component evaluation be repeated on a different set of office records to test the stability of the interpretations derived from the data.

3. Before undertaking the second recommendation, revise the profile component evaluation form as follows (see Figure 1):

- Reduce the number of worksheets to one and continue to use stick-on labels to identify charts for review.
- Ask the physician reviewer whether the physician writing in the record believes the patient has the condition listed on the label.
- Ask the physician reviewer whether the patient should receive the service, based on the evidence in the office record.
- If the patient should not receive the service, the physician reviewer should explain why.
- Finally, the physician reviewer should indicate whether he or she can find evidence in the office record of the service being provided.

Figure One
Proposed Revised Profile Component Evaluation Form

Assumes: Purpose of undertaking study is to understand threats to validity of the profiles, in particular, interpretations of the data that are discoverable from office record.

Label with
- patient identifiers
- profile/component
- service
- done by this doctor or other

I. Does the physician writing in the record believe the patient has the condition indicated on the label?

☐ Yes - Continue

☐ No - STOP

II. Based on the evidence in the record, should the patient receive the service indicated on the label?

☐ Yes - Continue
☐ Can't tell - Continue
☐ No

If No, then describe the patient, practice or other factors which explain why the service should not be given and then continue.

III. Can you locate evidence ***in this office record*** that the service was given?

☐ Yes

☐ No

Appendix 1

GENERAL

COMMENT	RESPONSE
Alabama: changes to CPT-4 in January 1992 might affect study.	It won't because the claims data end June 30, 1991.
Alabama: concerned about composition of comparison groups, i.e. a physician who devotes his practice to rheumatology may list his specialty as IM and thus he would be included in the IM comparison group when he shouldn't.	This problem is inevitable when dealing with claims data. We hope the mis-classification errors will cancel each other out. Both Maryland and Iowa have reported instances of physicians being sampled for the record review where they were either (1) not primary care physicians or (2) part of a "doc-in-the-box" group or a ER group.
Alabama: Concerned about using averages. This masks the range.	We have been sensitive to this issue. The focused profiles present percent of patients with 1 visit, 2-3 visits, etc.
Alabama: Did not like the word "primary" patient.	After trying many different wordings, primary patient still works best. We will retain.
Alabama: Looking at ambulatory quality issues is outside the scope of the study. The real issue is developing review tools.	Cannot develop review tools unless the tools have a purpose. In this case, the tools are to promote ambulatory care quality.
Sherrill, AL: Sample should exclude less serious cases since looking at ambulatory care.	Dr. Janssen from Iowa makes a similar suggestion for CHF (see below). We have incorporated the notion of stratification into the focused profiles to partially account for expected differences in service use.
Sherrill, AL: Many codes have been included under recommended care which may be inappropriate or rarely used in the care of the conditions.	We have reviewed the all codes. Our goal has been to be as inclusive as possible to capture all possibly appropriate use.
Maryland: Concerned about specialists acting as a major source of care. Would like an analysis of patients' utilization patterns to specialists or primary care physicians.	Table 3 of the focused profiles addresses the issue of service use by primary care patients of specialty care.
Maryland: DEMPAQ consultants should get own profiles before data are sent to rank and file. This will allow consultants to determine the strengths and weakness of current drafts and to offer advice about what to include in reports.	While we made profiles available to all liaison physicians prior to distributing the participant profiles, we received no feedback.

COMMENT	RESPONSE
Maryland: Concerned about handling of institutionalized patients.	Patients residing in nursing homes for greater than 3 months of the year were excluded.
Maryland: RU's sound too much like RBRVS which are offensive. Suggest Care Units (CUs) or Health Resources (HRs) or Medical Resources (MRs) or Medical Units (MUs).	We decided to retain RU's.
Maryland: Condition specific profiles should match AHCPRs or other guidelines.	AHCPR guidelines do not exist for the conditions we have chosen to review.
Maryland: Concerned about reporting info to an individual when the data are really about his or her group.	Some individuals received two profiles: an individual and a group profile. This was possible only for those individuals with both unique identifiers and group identifiers.
Powe: Most physicians will find the terms majority and non-majority rather awkward.	We have changed to primary.
Fisher: Don't know how to interpret majority and non-majority patients. Should test the sensitivity of findings to different definitions.	We changed the language back to primary.
Powe: Conditions chosen are important however absence of drug data in the claims severely limits their usefulness. Can't link the performance of BUN or creatinine to prescription of NSAIDs.	This is unfortunate; however the record review allows us to bridge the gap.
Luft: Diagnoses tend to be poorly coded.	The claims reliability study compares claims to data in the record. By the end of the study we should have a better sense of whether coding is a possible problem.
Fisher: Severity adjustments are not credible - they need more work.	We found it necessary to reduce the complexity of our severity stratifications in order to make them technically feasible. We chose a limited number of stratifications by condition and age.
Iezzoni: Should draw more on merged A/B data. Prepare profile of percent of physician' patients who had an ambulatory visit, x days, weeks, etc. after certain hospitalizations.	Table 2 of the focused profiles includes "Days between discharge and first ambulatory visit."

OFFICE PRACTICE PROFILE

COMMENT	RESPONSE
Alabama: not useful since about utilization.	We believe doctors will be very interested in the information provided by the office practice profile. Physicians do not usually have this kind of detailed information available to them.
Powe: Favor a three-row table for Table 1, with column headings: number of patients, percent of patients, number of visits, percent of visits.	This arrangement has been adopted.
Powe: Display average number of visits by majority patients before the table of distribution of ambulatory visits for majority patients.	Done.
Luft: Data on mean age of patient and age distribution is not of interest to physician. Perhaps a comparison to "similar" MDs would be more useful.	The facing page of each OPP table contains comparative data for the physician's specialty.
ACGs	
Powe: Statistics on ambulatory care groups need to be more detailed to be useful. In addition, some of the ACG modifiers are confusing (e.g. less than age 34) since the population is the over 65.	This summary has been replaced with a "Patient co-morbidity summary." The tables displaces number and percent of patients with no medical conditions, 1 medical condition, 2-3 conditions, etc.
Powe: For a clinician, a patient co-morbidity summary is more useful than the ambulatory care group summary.	See above.
Powe: ACG index confusing to most clinicians. Suggest presenting the index as percent greater or lower, e.g. total visits, x % greater than average based on case mix, etc.	We retained the ACG index and included text on the facing page to explain it.
Lohr: Use of ACGs in Medicare population problematic. Much care may be palliative. Physicians need to provide care which then shows up on a claim to evaluate for an ACG.	While some care is palliative, we are not sure we agree with the assertion that "much care" is palliative. Individuals in the Medicare population tend to have multiple conditions which do appear on the claim.
Iezzoni: Need to be careful with terminology. The ACGs are not a severity adjustment. They are a case mix adjustment.	Point well taken.

OFFICE PRACTICE PROFILE

COMMENT	RESPONSE
Iezzoni: Concerned that ACGs would be difficult to explain and not generally accepted by physicians.	This is a testable question. Liaison physicians seem to have accepted ACGs. When we receive feedback from participants we should have a clearer idea about the acceptability of ACGs
Iezzoni: A distribution of cases by organ system would be useful.	We considered this and rejected in favor of the co-morbid summary mentioned earlier. The co-morbid summary links directly to the ACGs which are presented later, while the organ system approach would stand alone.
Hospital stats	
Powe: Not sure it is useful to display % of patients discharged home to SNF.	We decided to retain this item.
Powe: Would like to see percentage of patients re-admitted within 30 days.	This can be considered for future versions of the profile.
Powe: Do not find inpatient visits per patient hospitalized useful because it will equal LOS.	We show both inpatient visits by "you" and by "others." This is potentially interesting in that it shows the extent to which "your" patients are being taken care of by others.
Resource Units	
Iezzoni: Use of resource units will be off-putting and imply a link to cost rather than to quality.	We have tried to be sensitive to this issue. Facing page texts explains RUs.
Powe: Need to be explicit when explaining resource units.	See above.
Powe: Need to explain what tests are considered chemistry, hematology, x-ray, etc.	Future versions of the profile could contain a fuller explanation of these services.

PREVENTION PROFILE

COMMENT	RESPONSE
Alabama: Concern about tests that are included that are not covered by Medicare, e.g. flu shots, mammography.	We share the concern. In fact we have chosen not to report back the preventive services profile data because the necessary data elements were not reliably in the data base.
Fisher: PSA should not be a recommended care item.	Agreed. We changed the designation to limited use.
Powe: Question inclusion of colonoscopy and prostate antigen as these are specific to clinical circumstances.	Objection taken care of by classifying the items as limited use.
Powe: Is a lipid study a single cholesterol determination?	Instead of referring to a lipid study, we create summary statistics for the individual components of a lipid study: total cholesterol, HDL cholesterol and triglycerides.

DIABETES PROFILE

COMMENT	RESPONSE
Severity	
Iezzoni: end-stage organ disease, e.g. peripheral vascular disease, renal disease, cardiac disease, ophthalmologic disease, prior hospitalization for amputation.	We elected to include the following in table 2: hospital admissions for complications: ketoacidosis, hyperosmolar coma, other coma, hypoglycemic coma, and amputation.
Recommended care	
Janssen, Iowa: Where are the FBS's, the 2 hr post prandial and the GTT's in the list of codes under blood glucose?	We had an experienced coder from Delmarva go over all the codes to check for problems.
Janssen, Iowa: Eliminate some of the urinalysis codes. These are hospital tests.	We argue that the physician should be given credit for any U/A test on this patient, no matter how specialized.
Hall, AL: Notes that recommended care items are somewhat difficult to define because they depend on the individual case. At a minimum, follow-up should occur every 3 months. The blood sugar and cholesterol should be checked every 4 months.	Our approach has been to present information about a minimal level of care.
Powe: "I do not think that hemoglobin A1C is recommended on a routine basis. The usefulness of this test has been debated for years in the literature."	We retain hgba1c.
Powe: Why are ophthalmological exams under general anesthesia included (92018-92019)?	
Limited use care--None listed for this diagnosis.	
Outcomes	
Powe: Outcomes, perhaps include admissions for cellulitis or ulcers of lower extremities.	We changed this table to include only selected conditions for which hospitalization might reflect poor ambulatory care.
Utilization	
Powe: Utilization - other specialists to track would be neurologists and general surgeons.	This specialists could be added to future versions of the profiles.

COPD

COMMENT	RESPONSE
Definition	
Lohr: Difficult to keep separate from asthma, CHF and acute bronchitis.	The classic definition of COPD includes only chronic bronchitis and emphysema. The other conditions can co-exist with COPD.
Severity adjustment	
Iezzoni: Home O_2 therapy, home respiratory therapy, prior admission for CHF, lung CA, pneumonia or peripheral vascular disease are all indicators of sicker patients.	Technically, stratification proved challenging because the analyses has to be run on all claims. We did stratify by CHF for COPD.
Recommended care	
Hall, AL: Recommended care includes yearly flu shot, blood gases (as required), pulmonary function studies.	We originally included all of these. The wording of Dr. Hall's letter, i.e. blood gases "as required," suggests that blood gases should be moved to the limited use column. This was done. In addition, we ultimately eliminated flu shot because of data quality difficulties.
Janssen, Iowa: Blood gases not done in the physician's office	The profile is intended to capture all, except inpatient, utilization, not just the services rendered in the physician's office. Thus blood gases drawn at a private laboratory will be captured.
Limited use care	
Janssen, Iowa: Suggests deleting a number of the chest x-ray codes as being done in the hospital rather than the office.	The profile is intended to capture all, except inpatient, utilization, not just the services rendered in the physician's office. Thus a chest xray performed at a facility will be captured.
Hall, AL: Limited use care includes CXRs, home O_2, and pulmonary rehab to improve reconditioning.	We include all except pulmonary rehabilitation. Pulmonary rehabilitation is not included in the profile because it presents special coding difficulties.
Powe: Are you going to include blood gases done in the hospital?	We will be able to include blood gases that appear on lab or institutional claims.
Powe: Blood gases and chest x-ray are a limited use item.	Chest x-ray already is a limited use item. We moved blood gases to the limited use category.
Outcomes	
Powe: Eliminate cough from outcomes list.	Good suggestion. Cough has been eliminated from the outcomes list.

ISCHEMIC HEART DISEASE

COMMENT	RESPONSE
Definition	
Iezzoni: Do not include patients with chest pain.	Chest pain as a condition listed on a claim probably represents a variety of conditions such as GI or lung problems. We did not include in the final version of the profiles.
Severity Adjustment	
Hall, AL: Severity adjustment - suggest adding hyperuricemia to list of severity adjustments for IHD.	As mentioned earlier, severity adjustment proved technically challenging. As a result we limited our severity adjustments to stratification by several diseases and age categories.
Powe: Question including chronic pulmonary disease and asthma as severity adjustments for this category.	See preceding note.
Fisher: AMI might be a good severity adjustor.	See above note.
Hall, AL: Suggest eliminating chronic pulmonary disease from the severity adjustment for IHD.	See above note.
Recommended Care	
Sherrill, AL: Recommended care - Some of the recommended care codes will never be used.	We attempted to be as inclusive as possible to make sure that credit was given, where appropriate, for care received.
Hall, AL: Recommended care, suggest adding cardiac enzymes, cholesterol and triglycerides.	We added total cholesterol and HDL cholesterol
Limited Use Care	
Powe: The following are limited use rather than recommended care items: catheterization, CABG, PTCA, ECG, exercise ECG, exercise thallium, and echocardiography.	We made these limited use items and eliminated altogether the CABG and PTCA.
Outcomes	
Hall, AL: Outcomes, add ventricular arrhythmias, CHF, aneurysms, AV block requiring pacemaker, pulmonary emboli, systemic emboli, and pericarditis.	We added ventricular arrhythmia and CHF.
Powe: Outcomes - why is stroke included? CHF should be added.	We eliminated stroke and added CHF.

CONGESTIVE HEART FAILURE

COMMENT	RESPONSE
Severity Adjustment	
Janssen, Iowa: Should focus on mild to moderate disease. The severe cases are treated in hospital.	This is a good point, however the claims data lack the sensitivity to identify those ambulatory cases that are mild versus severe.
Recommended care	
Janssen, Iowa: Delete selected CXR codes as being performed in the hospital, not in the office.	As noted earlier, we attempted to be inclusive in our coding.
Janssen, Iowa: Why is the stress test without the "interpretation and report" included.	The version of coding used in the profile accommodates all associated codes.
Hall, AL: Should add a renal profile and CBC.	We did not add.
Hall, AL: Should add contrast angiography, flow directed thermodilution catheter measurement and radionuclide ventriculography.	We added contrast angiography.
Powe: Recommended care, potassium should be done every 4 months. Flu shot every year.	The final version of the profiles did not include flu shots because of data problems.
Powe: Stress test done once in entire course of illness.	We moved this item to a limited use categorization.
Powe: echocardiography, ECG, chest x-ray are limited use, not recommended care because they are ordered for specific clinical indications.	These items were placed in the limited use category.
Limited use care: None listed in current version of profile	
Outcomes	
Hall, AL: Suggest adding ventricular arrhythmias, systemic emboli and cardiac transplantation.	We added arrhythmias and pulmonary embolism.
Powe: Outcomes - add hospitalization for stroke (from emboli) and arrhythmias.	We added pulmonary embolism and arrhythmias.

HYPERTENSION

COMMENT	RESPONSE
Severity	
Hall, AL: It might be more simple to name this draft Essential Hypertension rather than excluding all secondary causes of hypertension.	We decided to change the definition to exclude malignant HTN and CHF.
Powe: Exclude hypertensive renal disease and other secondary causes of hypertension.	The claims data do not permit us to determine the causes of a condition, therefore we cannot exclude cases with specific causes.
Iezzoni: End organ damage, e.g. renal disease.	Ultimately, we decided to severity stratify on age and by co-morbid IHD.
Recommended care	
Janssen, Iowa: Need to add to recommended care QA, EKG, CXR - PA and Lat, and lipid profile.	We added only total cholesterol and added it to limited use.
Hall, AL: The initial workup should include cholesterol, triglycerides, renal profile and EKG.	The claims data do not permit to determine at which point in the course of illness we have data. While the initial workup should include these items, we cannot limit our sample to individuals for whom the claim represents the initial workup.
Powe: Potassium measurements is limited use - only required if patient is on a diuretic.	Moved to limited use.
Limited use care	
Janssen, Iowa: Need to add to limited use care ambulatory BP monitoring.	We did not include ambulatory monitoring in the profile because of uncertainties about how the codes are used.
Hall, AL: Limited use care should include bio-feedback, home blood pressure, and possibly echocardiography to evaluate left ventricular size.	None of these items were included.
Outcomes	
Hall, AL: Should consider adding hypertensive crisis and hypotension secondary to treatment at outcomes.	Hypertensive crisis was judged too rare to include and hypotension secondary to treatment could not be identified from the claims.
Powe: Outcomes - exclude hypokalemia and retinopathy.	Retinopathy was excluded but hypokalemia retained.

OSTEOARTHRITIS

COMMENT	RESPONSE
Severity Adjustment	
Powe: Number of joints involved.	This depends upon consistent coding of the fifth digit modifiers. We determined that the data probably would not support this.
Iezzoni: Look for hip or compression fractures in previous year. Look for co-morbid conditions which affect osteoarthritis, e.g. malignancy, AMI, syncope.	Because we had only a single year's data available, we were unable to look for information from the prior year. Examining the data for co-morbid conditions was far too expensive for this project.
Recommended Care	
Janssen, Iowa: Add to recommended care, CBC, sed rate, arthritis profile, BUN, creatinine, liver function.	CBC, BUN, creatinine and liver function should be monitored for individuals on long term, high dose NSAIDs. Since we do not have drug data, we cannot determine which patients are on NSAIDs and which are not. An arthritis profile may vary from lab to lab. Sed rates can be used as markers of inflammation, however inflammation is unlikely if the condition is osteoarthritis.
Hall, AL: Under recommended care include, PT, transcutaneous nerve stimulation therapy, traction, arthroplasty, fusion (especially cervical fusion), total prosthetic replacement, angulation, osteotomy, laminectomy and discectomy.	The items Dr. Hall suggests are really limited use items. Physical therapy is already included. The others were excluded as being too infrequent.
Limited Use Care	
Hall, AL: For limited use should include steroid injection.	The quality and completeness of the data were suspect.
Janssen, Iowa: Add to limited use care, physical therapy, in particular massage, exercises and moist packs	Goroll recommends exercises, etc., and it would be appropriate to include the PT codes under limited use care. While these activities are "recommended" the performance of PT is a limited use. We included hot or cold packs and "exercise" in the profile. Exercise included functional activities, manipulation, and kinetic activities to increase coordination, strength or range of motion.

OSTEOARTHRITIS

COMMENT	RESPONSE
Outcomes	
Powe: Look at GI bleed as an outcome.	Excellent suggestion and good way to compensate for not having drug data. Of course, we can't directly link the bleed with drug use, but it might be a good proxy for problems with NSAIDs.
Powe: Must be careful in interpreting data on hip and knee replacement. Cannot determine the appropriateness of these operations using claims.	These data are not intended to be used to judge appropriateness. They are descriptive only.
Utilization	
Janssen, Iowa: Eliminate codes for arthrography of knee. These are hospital codes.	We eliminated arthroscopy altogether.

Appendix 2

Recommended Care

1) Primary care office (home) visit	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a minimum of two office visits within the study year, one in each half-year.	Recommended	According to the American Board of Family Practice, visits for diabetic patients in control should be scheduled every 3 to 6 months (ABFP, p. 36). This profile shows the percent of patients with diabetes who had a primary care visit at least once every 6 months.	None
Suggested Revisions	1. Change minimums with co-existing diseases. 2. Three to six months should be a maximum. Some patients require much more frequent visits to adjust and control.			

2) HgbA1c	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with at least one HgbA1c during the study year.	Recommended	Since the hemoglobin A1C test gives a good indication of the level of glucose control, it is recommended every 2 to 3 months (ABFP, p. 37) or every 6 months (ADA, p. 367).	None
Suggested Revisions	1. Might want to do twice per year (see Rationale).		1. Whether at 2-3 months or at 6 months depends on the severity of the diabetes and the brittleness of the disease.	

Structured Review of Face Validity: Diabetes Mellitus

3) Urinalysis	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with at least one urinalysis during the study year.	Recommended	Annual urinalysis is recommended to check for proteinuria (ABFP, p. 38; ADA, p. 367). Proteinuria is an early manifestation of diabetic nephropathy (Branch, p. 785).	None
Suggested Revisions				

4) Ophthalmology Visit	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with an eye exam during the study year.	Recommended	A complete eye and visual examination should be performed annually (ADA, p. 367). This examination is important to detect early signs of retinopathy (Harrison's, p. 1753) or cataract (Branch, p. 784).	None
Suggested Revisions	1. Eye exam during the study year for insulin-dependent, at 2 years otherwise.	1. Change to limited use.	1. Question need for annual eye exam in adult onset diabetes. 2. Should apply to patients with the severity stratification in Preventive Services for diabetes, hypertension, ischemic heart disease. Also for symptomatic patients. In addition, those on potentially oculo-toxic drugs - antimalarials, steroids, occupational exposure to lasers, etc. Not cost-effective to have annual visual exams for completely asymptomatic patients in any of the risk categories above.	1. Limited to severity stratification in Preventive Services for diabetes, hypertension, ischemic heart disease. Also those on potentially oculo-toxic drugs, occupational exposure to lasers, etc.

5) Influenza Immunization	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a flu shot during the study year.	Recommended	Influenza immunization is recommended by the Immunization Practices Advisory Committee from the Centers for Disease Control (MMWR 1991) and is particularly important for patients with chronic metabolic disease (Harrison's, p. 475).	None
Suggested Revisions			1. Exceptions: Demonstrated allergy to feathers and eggs, previous severe reaction to the immunization, and patients who refuse.	

6) Triglycerides	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a triglyceride level during the study year.	Recommended	Triglycerides should be tested annually (ADA, p. 367). Hypertriglyceridemia is common in diabetics (Harrison's, p. 1756).	None
Suggested Revisions	1. Reviewer does not think triglycerides should be a required component for all diabetics. Usually, if triglycerides are pathologically elevated, so are total and HDL cholesterol. Triglycerides are too sensitive to timing of last meal, whereas cholesterol and HDL are more reliable measures of the true "steady state" of person's lipid status.	1. Change to limited.	1. May not need to test annually (after initial screening and review of FH for familial patterns), but as patient's status, and past levels indicate. Good diet for diabetics will usually control hypertriglyceridemia.	1. Hypertriglyceridemia, positive FH, hyperlipoproteinuria.

7) Total Cholesterol	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a total cholesterol measurement during the study year.	Recommended	Total cholesterol should be tested annually (ADA, p. 367) because diabetics are prone to have arteriosclerosis and have an increased risk of myocardial infarction (Harrison's, p. 1733).	None
Suggested Revisions	1. After initial screening and FH, test as patient's status and past levels indicate.	1. Change to limited.	1. Consider total cholesterol/HDL, HDL/LDL ratios.	

8) HDL Cholesterol	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with an HDL cholesterol measurement during the study year.	Recommended	HDL cholesterol should be tested annually (ADA, p. 367). In diabetics the ratio of HDL to LDL is altered and the risk of arteriosclerotic disease is increased (Harrison's, p. 1753).	None
Suggested Revisions	1. After initial screening, test as patient's status and past levels indicate.	1. Change to limited.		

Limited Use

9) Serum Creatinine or Blood Urea Nitrogen	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with either a serum creatinine or urea nitrogen concentration during the study year.	Limited use	Serum creatinine or blood urea nitrogen concentrations should be measured if proteinuria is detected (ABFP, p. 38). Results of these tests reflect the degree of renal insufficiency.	Proteinuria or renal complications of diabetes Codes: 250.4
Suggested Revisions		1. Make recommended.	1. Serum creatinine and BUN should be measured on everyone.	1. Change to none.

10) Blood Glucose	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with at least one blood glucose during the study year.	Limited use	While many physicians use blood glucose to monitor patients with diabetes, guidelines from ABFP (p. 37) and ADA recommend HgbA1C for diabetes monitoring. In some cases, blood glucose tests may be a valuable adjunct to the HgbA1C test.	None
Suggested Revisions				

Congestive Heart Failure

Recommended Care

1) Primary care office (home) visit or visit with cardiologist	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a minimum of two office visits within the study year, one in each half year.	Recommended	The primary care visit for patients with CHF is important in monitoring drug doses, drug side effects, or symptoms of heart failure. Most primary care practitioners schedule office visits 2-4 months apart (Fleg et al., 1989). For this profile, 2 visits per year are considered minimal care.	None
Suggested Revisions	1. Why a cardiologist? Why can't family practitioner or general internist adequately manage this garden variety, grass-roots problem?			

2) Serum Potassium	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with at least one serum potassium during the study year.	Recommended	Close monitoring of serum potassium is necessary for patients who are receiving diuretics, as most patients with CHF do (Cohn, 1989).	None
Suggested Revisions			1. Also for those receiving digoxin or ACE inhibitors, potassium supplements.	

3) Influenza Immunization	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with an influenza immunization during the study year.	Recommended	Although annual influenza immunization is recommended for all persons 65 or older (Harrison's p. 474; Report of the US Preventive Services Task Force, p. 363), it is especially important for those with chronic diseases (i.e., heart disease), who have a high risk of pneumonia as a complication of influenza (Cassel et al., p. 320).	None
Suggested Revisions			1. Unless patient declines. 2. Exceptions: Demonstrated allergy to feathers and eggs, previous severe reaction to the immunization, patients who refuse.	

Limited Use

4)Echocardiography	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with an echocardiogram during the study year.	Limited use	Echocardiography (sonography) can provide accurate measurements of cardiac chambers, wall thickness, and valvular structures (ABFP, CHF Guide, p. 12). This technique should be limited to patients suspected of having cardiac valve disease, ventricular dysfunction, pericardial effusion, or a cardiac mass.	None
Suggested Revisions	1. If patient meets other criteria - valvular disease, pleural effusion.			1. Patients with valve disease, ventricular dysfunction, pericardial effusion, cardiac mass. 2. Patients with significant change of their condition.

5) ECG	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with an ECG during the study year.	Limited use	While ECGs do not confirm the presence or absence of heart failure, they provide clues to the etiology of the failure (ABFP, CHF Guide, p. 11).	None
Suggested Revisions	1. All patients should have ECG once per year minimum.	1. Make recommended. 2. Make recommended.	1. Need to assess rhythm change, PR interval change, and STT wave change at least yearly.	1. All those without very stable disease. (Those with very stable disease do not need "routine" ECG.)

6) Chest x-ray	**A. Component**	**B. Attribute**	**C. Rationale**	**D. Severity Stratification**
Proposed	Patients with a chest x-ray during the study year.	Limited use	The chest radiograph provides an indication of underlying cardiac abnormality and information regarding the status of the failure itself (ABFP, CHF Guide, p. 11). While useful, the chest x-ray is not a very sensitive screening test for heart failure (Branch, p. 129). The history and signs detected at physical exam are more important for monitoring and managing the patient with CHF.	Bronchitis, COPD, pneumonia **Rationale:** Chest x-rays are helpful in evaluating patients with CHF and one of these conditions. **Codes:** 480.xx - 486.xx, 490.xx, 491.xx - 492.xx, 496.xx
Suggested Revisions		1. If CHF unstable.	1. Unless signs of failure present, not sensitive enough to screen - needs revision.	1. Add change in breathing pattern and/or elevated S.O.B.

7) Contrast Angiography	**A. Component**	**B. Attribute**	**C. Rationale**	**D. Severity Stratification**
Proposed	Patients with a contrast angiogram during the study year.	Limited use	Injection of radiographic contrast agent into coronary arteries or left ventricle permits assessment of coronary artery stenosis or ventricular ejection fraction (Harrison's, pp. 872-4). This technique is useful only for patients known or suspected to have IHD in addition to CHF.	Ischemic heart disease **Rationale:** Management of the patient with CHF depends on the presence of additional risk factors (ABFP, CAD Guide, p.13). **Codes:** 410.xx - 414.xx
Suggested Revisions	1. Only if the patient would be candidate for angioplasty or CABG.		1. Agree if patient is surgical candidate. (Appropriate LV function.) 2. Comment: Contrast angiography should be performed when there is an intention to act upon the findings, if it is indicated. (Consider Alzheimers patient in nursing home with suspected IHD and CHF.)	

8) Exercise Stress Test (ECG)	**A. Component**	**B. Attribute**	**C. Rationale**	**D. Severity Stratification**
Proposed	Patients with a cardiovascular stress test during the study year.	Limited use	Stress testing provides information about ventricular function and myocardial ischemia (Harrison's, p. 966). Exercise testing may help identify the cause of heart failure (ABFP, CHF Guide, p. 12). Exercise tolerance is an important tool to evaluate the efficacy of treatment (Cohn, 1988) and to identify a safe range for heart rate during exercise (Merck, pp. 280-1). However, exercise testing is contraindicated for patients with unstable angina, probable AMI, severe aortic stenosis or hypertension, uncontrolled arrhythmias (Branch, p. 57) or for those who are asymptomatic (Branch, p. 63).	Ischemic heart disease **Rationale:** See contrast angiography. **Codes:** 410.xx - 414.xx
Suggested Revisions			1. Add: Patients showing "T" and "ST" changes without one for comparison that was done prior, should have a stress test or thallium or persantine stress done. Stress tests are not contraindicated for some asymptomatic patients with EKG changes. 2. Also contraindicated for those in whom the diagnosis of anginal ischemia is conclusive on clinical grounds, and for whom more aggressive, invasive intervention is not contemplated.	1. Add changes in C.1.

Recommended Care

1) Primary care office (home) visit	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a minimum of two office visits within the study year, one in each half year.	Recommended	The primary care visit is important in monitoring the course of a progressive disease. A minimum of 3 visits per year is recommended by the American Board of Family Practice (ABFP, COPD Guide, p. 35). For this profile, 2 visits per year are considered minimal care.	None
Suggested Revisions				

2) Influenza Immunization	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a flu shot in the study year.	Recommended	Influenza immunization is particularly important for patients with COPD to help prevent exacerbation of disease. A yearly immunization is recommended by several sources (Harrison's, p. 475; Branch, p. 326; Report of the US Preventive Services Task Force, p. 363).	None
Suggested Revisions				

Limited Use

3) Spirometry	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a spirometry exam during the study year.	Limited use	Spirometry provides a measure of a patient's vital capacity and lung functioning. The ABFP COPD Guide recommends that spirometry be performed at every routine visit (p. 35). Spirometry has been placed in the limited use column to reflect that patients with mild disease may not require frequent or even yearly spirometry.	Any diagnosis of pneumonia, lung cancer, CHF, influenza, asthma, or bronchitis **Rationale:** Patients with these conditions in the study year are more likely to have compromised breathing capacity. **Codes:** 162-163.9, 480-490, 491-493.9,496
Suggested Revisions				

4) Chest X-ray	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a chest x-ray during the study year.	Limited use	A chest x-ray is important in patients with exacerbations of disease to check for pneumonia. Chest x-rays can also assist in making a diagnosis of CHF for patients with COPD. Chest x-rays also can provide information about hyperinflation and changes in vascular markings (Scientific American, Chronic Obstructive Disease of the Lung, p. 11). The ABFP COPD Guide (p. 35) suggests that yearly chest x-rays may be warranted by the patient's clinical condition.	Any diagnosis of pneumonia, lung cancer, CHF, influenza, asthma, or bronchitis **Rationale:** See spirometry above. **Codes:** 162-163.9, 480-490, 491-493.9, 496
Suggested Revisions	1. Only with a change in baseline condition.		1. Annual CXR not required with stable disease. Also not required for every exacerbation of disease, only when there is a failure to promptly respond to adequate treatment. Annual CXR not mandated in bedridden institutionalized patients where the only study would be a portable chest film.	1. Progression of symptoms. 2. May be warranted yearly depending on patient's clinical condition.

5) Home Oxygen Therapy	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients on home oxygen therapy during the study year.	Limited use	Patients with severe COPD ($pO_2 < 60$ mmHg) may benefit from home O_2 therapy (Goroll, p. 218), but the need should be documented. Physicians with a higher proportion of patients on home oxygen therapy than the statewide average may have a patient population that is more severely ill.	Any diagnosis of pneumonia, lung cancer, CHF, influenza, asthma, or bronchitis **Rationale:** See spirometry above. **Codes:** 162-163.9, 480-490, 491-493.9, 496
Suggested Revisions	1. Liberalize criteria - not all COPD patients need home O_2.		1. Most insurance will not cover home O_2 unless PO_2 <55. Maybe needs revision to <PO_2 55. 2. May be very helpful to prevent cognitive impairment in permanent home care even if PO_2 >60. 3. Any patient with persistent PO_2 <55 or <60 with RHF or cor pulmonale should be on home O_2.	1. Spirometry done and PO_2 <60 mm Hg.

6) Blood Gases	**A. Component**	**B. Attribute**	**C. Rationale**	**D. Severity Stratification**
Proposed	Patients with an outpatient blood gas determination during the study year.	Limited use	Blood gases are important to measure oxygenation of the blood and CO_2 retention, factors which may affect the treatment plan. Yearly measurement of blood gases is recommended by the American Board of Family Practice (ABFP, COPD Guide, p. 35). Blood gases have been placed in the limited use column to reflect the fact that patients with mild disease may not need a yearly blood gas determination. (Note: Blood gas measurements done in hospital are excluded.)	Home O_2 **Rationale:** Patients whose COPD is severe enough to require O_2 therapy should have blood gases measured at least yearly (ABFP, COPD Guide, p.35).
Suggested Revisions	1. Use of pulse oximeter may be substituted. 2. Should be ABG with or without oxygen saturation.		1. Some patients with significant disease, but stable, may not need annual ABG if their lung disease is such that their oxygenation is adequate for their lifestyle (unless using O_2).	ABG/O_2 saturation.

Preventive Services

Recommended Care

1) Influenza immunization	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with an influenza immunization during the study year.	Recommended	Annual influenza immunization is recommended for all persons 65 or older (Harrison's p. 474); Report of the US Preventive Services Task Force, p. 363; Merck, p. 917).	None
Suggested Revisions			1. Patient may refuse and the refusal may not always be documented. 2. Exceptions: a. demonstrated allergy to feathers and eggs b. previous severe reaction to the immunization c. patients who refuse immunization.	1. See C.2.

2) Mammogram (female only)	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Female patients less than 75 years old with a mammogram during the study year.	Recommended	An annual mammogram is recommended for women aged ≤75 (Report of the US Preventive Services Task Force, p. 363; Merck, p. 917).	None
Suggested Revisions			1. Should include all women, even ≥75 yrs. 2. Some studies recommend annual mammograms after 75 years. Make "limited use" apply to mammograms after 75, with stratification.	

Limited Use

3) Total cholesterol	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a total cholesterol during the study year.	Limited Use	The US Preventive Services Task Force (p. liv) recommends testing of serum cholesterol every year. The American Colletge of Physicians (Hayward et al., p. 769) and the ABFP (ABFP, p. 37) recommend frequent or appropriately times cholesterol tests for patients with IHD. Frequency of testing depends on past cholesterol levels and other factors in the patient's history and status.	Diabetes mellitus ICD9: 250-250.91 Hypertension ICD9: 401-401.9 Ischemic heart disease ICD9: 410.xx-441.xx
Suggested Revisions	1. Lipid profile more appropriate for screening. 2. Agree, but if total is >240, should have HDL and LDL. 3. Include HDL testing as well, and make the component a total cholesterol/HDL ratio.	1. Change to recommended. 2. Change to recommended. 3. Change to recommended.	1. Also depends upon prior treatment and age (would not need yearly cholesterol on someone >80).	1. None - every 5 years on all patients. 2. Change to none. 3. Add hyperlipoprotenemia, hypercholesterolemia, peripheral vascular disease, ASHD, etc. and cerebrovascular disease, arteriosclerosis.

4) Pap smear (female only)	**A. Component**	**B. Attribute**	**C. Rationale**	**D. Severity Stratification**
Proposed	Female patients with a pap smear during the study year.	Limited Use	Recommended for women who have not had previous documented screening in which smears have been consistently negative (US Preventive Services Task Force, p. liv).	None
Suggested Revisions	1. Yearly not needed for women after hysterectomy or for women with 2 negatives. 2. Add: If patient has an intact uterus and/or cervix.	1. Change to recommended with modifications. 2. Change to recommended.	1. Would screen every 2-3 years in low-risk category; more often in others, based on clinical situation. 2. Recommended for all women >65 who have intact uterus and/or cervix.	1. Make exceptions for those patients with hysterectomies (less frequent paps).

5) Colonoscopy	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a colonoscopy during the study year.	Limited Use	Colonoscopy is recommended for persons with a family history of polyposis coli or cancer family syndrome (US Preventive Services Task Force, p. liv).	Diagnosis of polyposis coli ICD9: 153-153.9, 211.3
Suggested Revisions	1. Need to take into account colonoscopy done every 1-3 years after positive family history.		1. Recommended for persons with positive family history, but not every year. 2. Add personal history of polyps seen or symptoms with abnormal barium enema. 3. Also patients with prior malignant polyps or prior colonic malignancy. 4. Add history of previous colonic malignancy. 5. Add patient with rectal bleeding with or without positive stools should be same indication as sigmoidoscopy. (Colonoscopy and sigmoidoscopy should be one preventive service, with specific test used as judgment decision.)	1. Previous colon cancer or polyps. 2. See C.3.

6) Sigmoidoscopy	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Patients with a sigmoidoscopy during the study year.	Limited Use	Sigmoidoscopy is recommended for persons with a history of endometrial, ovarian, breast, or colorectal cancer, inflammatory bowel disease or adenomatous polyps, or those with first degree relatives who have colorectal cancer (US Preventive Services Task Force, p. liv).	Diagnosis of cancer ICD9: 153-153.9, 174-174.9, 182-183.9, 555-555.9, 558.9
Suggested Revisions		1. Make recommended. 2. Should be recommended for screening.	1. Sigmoidoscopy can be done every 3-5 years if two previous negative studies. First-degree relatives with colorectal cancer need colonoscopy, not sigmoidoscopy. 2. Screening every 3-5 years (recommended). 3. Also include unexplained weight loss, heme positive stools, and change in bowel habits. 4. Colonoscopy and sigmoidoscopy should be one preventive service, with specific test used as judgment decision. 5. Include those with rectal bleeding, positive occult blood, and periodic screening for colorectal cancer, i.e., at 3-5 years after consecutive negative exams.	1. None 2. See C.3. 3. Include rectal bleeding.

7) Prostate-specific antigen test	A. Component	B. Attribute	C. Rationale	D. Severity Stratification
Proposed	Male patients with a prostate-specific antigen test.	Limited Use	Although the PSA test is not useful for screening for prostate cancer, when combined with other diagnostic techniques it may be helpful for identification or follow-up of the disease (Harrison's, p. 1631).	Prostate cancer
Suggested Revisions	1. Agree if not used as a screen.	1. Recommend for all males over 65 - helpful test in screening for prostate cancer. 2. Change to recommended.	1. a. All male patients over 65 to have PSA. b. If over 40, to have sonogram. c. If questionable area seen on sonogram, should have multiple biopsies. 2. Technology has changed, and PSA should be recommended annual procedure for all males.	1. BPH to monitor any potential change and schedule Bx or ultrasound. 2. Add those with BPH, nodules, or change in consistency. 3. All males ≥65. 4. Include BPH.

Appendix 3

II. PATIENT-SPECIFIC CASE REVIEW

A. Recommended Care: Care Given

A recommended care item is one that should be given to the patient.

I. Does the patient have the condition indicated on the label?

☐ No → STOP.

☐ Yes → Proceed to Section II.

II. If the label indicates that the care was given:

- Are there any factors in this particular case that suggest the care was inappropriate?

  ☐ No, the care was appropriate for this patient → STOP.

  ☐ No, no such factors are documented in the medical record → STOP.

  ☐ Yes → Continue below.

- Describe the factors or circumstances that suggest the care was inappropriate:

  ☐ **Patient factors** (please explain):

  ☐ **Practice factors** (please explain):

  ☐ **Extenuating circumstances** (please explain):

- Comments:

II. PATIENT-SPECIFIC CASE REVIEW

A. Recommended Care: Care NOT Given

A recommended care item is one that should be given to the patient.

I. Does the patient have the condition indicated on the label?

☐ No → STOP.

☐ Yes → Proceed to Section II.

II. If the label indicates that the care was not given:

- Are there any factors in this particular case that justify the absence of care?

  ☐ No; justification is not documented in the record → STOP.

  ☐ Yes → Continue below.

- Describe the factors or circumstances that explain the absence of care:

  ☐ **Patient factors** (please explain):

  ☐ **Practice factors** (please explain):

  ☐ **Extenuating circumstances** (please explain):

- Comments:

II. PATIENT-SPECIFIC CASE REVIEW

B. Limited Use Care: Care Given

Use of care items in the limited use category depends on the patient's condition and the physician's discretion.

I. Does the patient have the condition indicated on the label?

☐ No → STOP.

☐ Yes → Proceed to Section II.

II. If the label indicates that the care was given:

- Do the facts in this particular case justify the item's use?

☐ No, justification for the use of care is not documented in the record → STOP.

☐ Yes → Continue below. ↙

☐ No; care is inappropriate → Complete next item below. ↙

· Are there diagnoses or procedures which justify the use of this item? ☐ No → STOP. ☐ Yes → Complete next item below. ↙	• What factors suggest that the care was inappropriate? ☐ **Patient factors** (please explain):
• List the diagnoses or procedures below:	☐ **Practice factors** (please explain): ☐ **Extenuating circumstances** (please explain)

- Comments:

II. PATIENT-SPECIFIC CASE REVIEW

B. Limited Use Care: Care NOT Given

Use of care items in the limited use category depends on the patient's condition and the physician's discretion.

I. Does the patient have the condition indicated on the label?

☐ No → STOP.

☐ Yes → Continue.

II. If the label indicates that the care was not given:

- Do the facts in this particular case justify the absence of care?

☐ No, justification for the absence of care is not documented in the record → STOP.

☐ Yes → Continue below.

- Describe the factors or circumstances that **explain** the absence of care:

☐ **Patient factors** (please explain):

☐ **Practice factors** (please explain):

☐ **Extenuating circumstances** (please explain):

- Comments:

Appendix 4

Profile Component Evaluation
CHF

Office Visit

	Patient Needed Service?* Yes	No	Can't tell	
In Claim	28	0	0	28
Not in Claim	1**	0	0	1
	27	0	2	29

How often a claim indicates good quality: 100%
How often no claim indicates good quality: 0%

* Based on evidence in the office record
** Patient appropriately received services according to the office record - at VA

Influenza Immunization

	Patient Needed Service?* Yes	No	Can't tell	
In Claim	2	0	0	2
Not in Claim	3**	1	23	27
	5	1	23	29

How often a claim indicates good quality: 100%
How often no claim indicates good quality: 4%

* Based on evidence in the office record
** Two patients appropriately received services through Public Health, 1 patient at the VA

Potassium

	Patient Needed Service?* Yes	No	Can't tell	
In Claim	2	0	2	2
Not in Claim	18**	2	7	27
	20	2	9	29

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 8%

* Based on evidence in the office record
** Although no claims were located, 18 patients appropriately received services according to the office record. This is due to a coding problem during the programming for the profile component evaluation

Angiography

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	1	0	0	1
Not in Claim	0	24	4	28
	1	24	3	29

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 86%

* Based on evidence in the office record

ECG

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	20	0	3	23
Not in Claim	3**	2	1	6
	23	2	4	29

How often a claim indicates good quality:, 87%
How often no claim indicates good quality: 33%

* Based on evidence in the office record
** Three patients appropriately received services at the VA or hospital

Chest X-ray

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	22	1	2	25
Not in Claim	2**	1	1	4
	24	2	3	29

How often a claim indicates good quality:, 88%
How often no claim indicates good quality: 25%

* Based on evidence in the office record
** Two patients appropriately received services at the VA or in the hospital

Echocardiography

	Patient Needed Service?* Yes	No	Can't tell	
In Claim	10	0	2	12
Not in Claim	3**	8	6	17
	13	8	8	29

How often a claim indicates good quality: 83%
How often no claim indicates good quality: 47%

* Based on evidence in the office record
** Three patients appropriately received services at the hospital or from the VA

Stress Test

	Patient Needed Service?* Yes	No	Can't tell	
In Claim	1	0	1	2
Not in Claim	1**	18	8	27
	2	18	9	29

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 67%

* Based on evidence in the office record
** This patient appropriately received services in the hospital

Profile Component Evaluation
COPD

Office Visit

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	28	0	0	28
Not in Claim	0	0	0	0
	28	0	0	28

How often a claim indicates good quality: 100%

* Based on evidence in office record

Influenza Immunization

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	8	0	0	8
Not in Claim	5**	1	14	20
	13	1	14	28

How often a claim indicates good quality: 100%
How often no claim indicates good quality: 5%

* Based on evidence in office record
** Five patients appropriately received services, probably through Public Health

Blood Gases

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	6	0	1	7
Not in Claim	8**	9	4	21
	14	9	5	28

How often a claim indicates good quality: 86%
How often no claim indicates good quality: 43%

* Based on evidence in office record
** These patients appropriately received services in the hospital

Chest X-Ray

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	20	1	1	22
Not in Claim	0	5	1	6
	20	6	2	28

How often a claim indicates good quality: 91%
How often no claim indicates good quality: 83%

* Based on evidence in office record

Spirometry

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	11	0	0	11
Not in Claim	1**	8	8	17
	12	8	8	28

How often a claim indicates good quality: 100%
How often no claim indicates good quality: 47%

* Based on evidence in office record
** According to the office record, this patient needed services and didn't get them

Home O2

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	4	1	0	5
Not in Claim	0	17	6	23
	4	18	6	28

How often a claim indicates good quality: 80%
How often no claim indicates good quality: 74%

* Based on evidence in office record

Profile Component Evaluation
Diabetes

Office Visit

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	29	0	0	29
Not in Claim	1**	0	0	1
	30	0	0	30

How often claim indicates good quality: 100%
How often no claim indicates good quality: 0%

* Based on evidence in the office record
** This patient appropriately received services through Medicaid

Influenza Immunization

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	1	0	0	1
Not in Claim	12**	3	14	29
	13	3	14	30

How often a claim indicates good quality:, 100%
Predictive value of no claim, 10%

* Based on evidence in the office record
** These patients probably received services from Public Health, according to the office record

HgbA1c

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	7	0	0	7
Not in Claim	2**	0	21	23
	9	0	21	30

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 0%

* Based on evidence in the office record
** Although no claims were located, 2 patients appropriately received services according to the office record. This probably represents the codi problem that occurred during the programming of the profile component evaluation.

HDL Cholesterol

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	5	0	0	5
Not in Claim	8**	3	14	25
	13	3	14	30

How often a claim indicates good quality: 100%
How often no claim indicates good quality: 12%

* Based on evidence in the office record

** Although no claims were located, 8 patients appropriately received services according to the office record. This probably represents the coding problem that occurred during the programming of the profile component evaluation.

Total Cholesterol

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	7	0	0	7
Not in Claim	10**	4	9	23
	17	4	9	30

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 17%

* Based on evidence in the office record

** Although no claims were located, 10 patients appropriately received services according to the office record. This probably represents the coding problem that occurred during the programming of the profile component evaluation.

Urinalysis

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	18	0	0	18
Not in Claim	2**	0	10	12
	20	0	10	30

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 0%

* Based on evidence in the office record

** The charges for these two patients were probably included in the office charge, according to the physician's review of the medical record

Profile Component Evaluation
Diabetes

Office Visit

Patient Needed Service?*

	Yes	No	Can't tell	
In Claim	29	0	0	29
Not in Claim	1**	0	0	1
	30	0	0	30

How often claim indicates good quality: 100%
How often no claim indicates good quality: 0%

* Based on evidence in the office record
** This patient appropriately received services through Medicaid

Influenza Immunization

Patient Needed Service?*

	Yes	No	Can't tell	
In Claim	1	0	0	1
Not in Claim	12**	3	14	29
	13	3	14	30

How often a claim indicates good quality:, 100%
Predictive value of no claim, 10%

* Based on evidence in the office record
** These patients probably received services from Public Health, according to the office record

HgbA1c

Patient Needed Service?*

	Yes	No	Can't tell	
In Claim	7	0	0	7
Not in Claim	2**	0	21	23
	9	0	21	30

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 0%

* Based on evidence in the office record
** Although no claims were located, 2 patients appropriately received services according to the office record. This probably represents the coding problem that occurred during the programming of the profile component evaluation.

HDL Cholesterol

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	5	0	0	5
Not in Claim	8**	3	14	25
	13	3	14	30

How often a claim indicates good quality: 100%
How often no claim indicates good quality: 12%

* Based on evidence in the office record

** Although no claims were located, 8 patients appropriately received services according to the office record. This probably represents the coding problem that occurred during the programming of the profile component evaluation.

Total Cholesterol

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	7	0	0	7
Not in Claim	10**	4	9	23
	17	4	9	30

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 17%

* Based on evidence in the office record

** Although no claims were located, 10 patients appropriately received services according to the office record. This probably represents the coding problem that occurred during the programming of the profile component evaluation.

Urinalysis

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	18	0	0	18
Not in Claim	2**	0	10	12
	20	0	10	30

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 0%

* Based on evidence in the office record

** The charges for these two patients were probably included in the office charge, according to the physician's review of the medical record

Ophthalmology Visit

Patient Needed Service?*

	Yes	No	Can't tell	
In Claim	12	0	0	12
Not in Claim	1*	2	15	18
	13	2	15	30

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 20%

* Based on evidence in the office record
** This patient appropriately received services through Medicaid

Triglycerides

Patient Needed Service?*

	Yes	No	Can't tell	
In Claim	8	0	0	8
Not in Claim	11**	3	8	22
	19	3	8	30

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 14%

* Based on evidence in the office record
** Although no claims were located, 11 patient** appropriately received services according to the office record. This probably represents the coding problem that occurred during the programming of the profile component evaluation.

Blood Glucose

Patient Needed Service?*

	Yes	No	Can't tell	
In Claim	22	0	0	22
Not in Claim	5**	1	2	8
	27	1	2	30

How often a claim indicates good quality: 100%
How often no claim indicates good quality: 13%

* Based on evidence in the office record
** Although no claims were located, 5 patients appropriately received services according to the office record. This probably represents the coding problem that occurred during the programming of the profile component evaluation.

Profile Component Evaluation

Prevention

Flu Shot

	Patient Needed Service?[*]			
	Yes	No	Can't tell	
In Claim	10	0	0	10
Not in Claim	6[**]	3	31	40
	16	3	31	50

How often a claim indicates good quality: 100%
How often no claim indicates good quality: 8%

[*] Based on evidence in the office record

[**] Six patients appropriately received services according to the office record - apparently at the office, perhaps charge included in global office visit charge

Mammogram

	Patient Needed Service?[*]			
	Yes	No	Can't tell	
In Claim	0	0	0	0
Not in Claim	2[**]	2	14	18
	2	2	14	18

How often a claim indicates good quality: NA
How often no claim indicates good quality: 11%

[*] Based on evidence in the office record

[**] Two patients appropriately received services, according to the record. No mention where.

Sigmoidoscopy

	Patient Needed Service?[*]			
	Yes	No	Can't tell	
In Claim	2	0	0	2
Not in Claim	0	18	30	48
	2	18	30	50

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 38%

[*] Based on evidence in the office record

Pap Smear

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	2	0	0	2
Not in Claim	0	7	24	31
	2	7	24	33

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 23%

* Based on evidence in the office record

Colonoscopy

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	1	0	0	1
Not in Claim	0	19	29	48
	1	19	29	49

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 40%

* Based on evidence in the office record

PSA

	Patient Needed Service?*			
	Yes	No	Can't tell	
In Claim	0	0	0	0
Not in Claim	0	7	10	17
	0	7	10	17

How often a claim indicates good quality:, NA
How often no claim indicates good quality: 41%

* Based on evidence in the office record

Total Cholesterol

	Patient Needed Service?* Yes	 No	 Can't tell	
In Claim	5	0	0	5
Not in Claim	18**	9	18	45
	23	9	18	50

How often a claim indicates good quality:, 100%
How often no claim indicates good quality: 20%

---

* Based on evidence in the office record

** Although no claims were located, 18 patients appropriately received services according to the office record. This probably represents the coding problem that occurred during the programming of the profile component evaluation.

Chapter 11
The Record Review Instrument

Abstract

DEMPAQ Record Review System

We developed a computerized record review system around typical activities performed in the course of an office visit which we called *"functions,"* for example, the appropriate prescription of a drug for a patient and then monitoring the patient for adverse side effects. Functions are performed for specific *"clinical items,"* e.g., for a specific drug. Summary measures of quality for key aspects of the clinical processes related to functions, e.g., the rate of drug prescriptions that were clearly appropriate for a given drug are known as *"indicators."* Unique *"criteria"* have been developed for each clinical item/indicator pair, such as a list of the appropriate indications for a specific drug. We also developed a rating for the adequacy of documentation in the record.

The record review system was programmed in Paradox, a common database language. The system consisted of a group of modules, one for each function. Within each module, a series of screens displayed the questions posed by the indicators and presented the criteria to evaluate the physician's office record in a box at the center of the screen.

Sampling

Three hundred randomly selected physicians in each state were invited to participate in the DEMPAQ project. A list of 25 randomly selected patients accompanied each sampled physician's letter. Ninety-one physicians agreed to participate in Maryland, 100 in Alabama (only 92 physicians were actually reviewed, however) and 53 in Iowa.

Conducting the Review

The steps in conducting the review included inviting the physicians to participate and receiving their records, preparing the office record for review, and abstracting the record. The project also instituted strict data quality control methods to ensure the integrity of the collected data.

Scoring and Rate Calculation

At the completion of the review, each indicator for each case was assigned one of five scores: met; acceptable alternative; not applicable; not reviewable; and flag. The findings were then aggregated to create performance rates for all cases of an individual physician. Performance rates were prepared at the clinical item, indicator and function levels for the individual physician with comparisons to the physician's peer specialty group.

Reports

Two types of report books have been developed. The report books for the liaison physicians include data on the reliability of the data abstraction and the validity of the criteria, as well as data on the percent acceptable performance for all physician participants in a state. Report books for individual physicians include clinical item reports, indicator summary reports, a function summary report and a record documentation report. All reports present data for the individual physician with comparisons to his or her peer group. This chapter describes only the reports used in the group meetings with liaison physicians. The individual participant physician reports will be described as part of the report on the Feedback and Education Modification to the Original DEMPAQ Contract.

Suggestions for the Future

The chapter concludes with specific suggestions on what needs to be done prior to using DRRS again.

Introduction

This chapter is about the DEMPAQ Record Review System (DRRS), its design, construction and implementation. The record review portion of the DEMPAQ project was intended to build upon the clinical detail that may be present in a physician's office record. A physician's impressions, hypotheses, and decisions, the results of laboratory and imaging studies, as well as the drugs and dosages prescribed, may be found in an outpatient chart. A record review system can tap this potential wealth of information.

Physician's office records also have their disadvantage. A frequent criticism of the record is the lack of documentation in the record by physicians. In a study published in 1981, Romm and Putnam looked at the frequency of missing data in the ambulatory patients record (Romm and Putnam, 1981). They compared transcripts of actual visits to information recorded in the chart. The patient's chief presenting complaint was absent from the record in only 8% of the cases; the physician's diagnosis absent in 7% of the cases. Tests and planned therapy were missing 13% and 15% of the time, respectively.

The following sections describe the structure of the review, the design of the computerized record review system, physician and patient sampling for the review, conducting the review, scoring and reporting the results of the review.

Structure of the Review

The DEMPAQ record review takes an in-depth look at patterns of care given to elderly Medicare patients in physician offices. The following outlines the framework of the review system.

Functions

The DEMPAQ record review tool is built around clinical activities that are commonly performed in the course of an office visit; these activities are called functions. The following functions are reviewed:

- Making a ***diagnosis***;
- Prescribing a ***drug***;
- Responding to an abnormal ***test***;
- Ordering or performing a ***procedure***;
- Performing ***preventive*** or ***screening*** services; and
- ***Follow-up*** of serious health threats.

Rather than structuring the review solely around diagnoses as is common with many other performance measurements, the DEMPAQ record review covers a wide variety of ambulatory health care activities. By measuring performance from several perspectives, the quality of health care is validated from different viewpoints. The structure also enables the review to proceed despite the variety of styles of physician record keeping. For example, physicians who do not usually document diagnoses, but only record their decisions related to drugs, can still be reviewed by DEMPAQ for drug-related care. This feature makes the DEMPAQ record review unique.

Clinical Items

Examples of functions, that is, specific drugs or tests, are called clinical items. For instance, drug clinical items include, among others: ACE inhibitors; calcium channel blockers and corticosteroids; test clinical items include blood pressure, creatinine and blood glucose. The criteria for selecting clinical items considered: how often the item occurred in the elderly population, and whether it contributed to the diversity of clinical items already selected for review.

New lists of clinical items can be easily formulated to address the needs of particular populations. Diagnoses for pediatric populations might include asthma or "well child care." Low back pain or pregnancy might be included in a review of young to middle aged adults.

Indicators

We developed summary measures of quality for key aspects of the clinical processes and activities that influence patient outcomes. Several measures, called indicators, were developed for each function. Each indicator is a rate of performance for a process of patient care believed to be important in securing a good patient outcome. For example, the indicators for the Drug function are:

- Drug dose and frequency appear in the record;
- Drug is prescribed for an appropriate condition;
- There are no condition contraindications in the record, or if present, they are justified;
- There are no drug contraindications in the record, or if present, they are justified; and
- The patient is appropriately monitored for adverse effects of the medication.

Criteria

Unique criteria have been developed for the review of each clinical item/indicator combination. The criteria answer the question "Was performance appropriate for this indicator and for this item?" For example, the actions which represent appropriate monitoring for a diagnosis of diabetes differ from the appropriate monitoring actions for ischemic heart disease.

As described in Chapter 3, the criteria were developed with the assistance of major primary care medical societies in each of the three study states. Representatives of each organization reviewed the proposed criteria in an iterative procedure involving mailings, phone calls and group meetings. The same physicians took part in the evaluation of the performance measures (see Chapter 14). Subsequently they appraised the results of the review.

DRRS Computer System

We developed an interactive computer program to facilitate conducting the record review. The program was written in Paradox, a common database software package (Paradox, 1990). The program's framework was built around the indicators described earlier. Paradox dictionaries contained the specific review criteria for each function.

A computerized system had three important advantages over a paper system from our perspective. First, and perhaps most importantly, a computerized review instrument can easily accommodate complex branching in the review structure. For example, the first choice made by a reviewer is whether the reviewed item is a drug, test or diagnosis. Depending on the answer to this question, the reviewer is next shown a list of appropriate clinical items for the function and asked to identify the item being reviewed. If the function is drugs or diagnoses, the reviewer is queried about the status of the drug or diagnosis new or old? This branching structure would require considerable flipping back and forth of a paper questionnaire, but is accomplished seamlessly by a computer system.

A second advantage of a computer system is that it can be easily adapted to changes in medical practice and standards. The structure of our system, a generic program accommodating all the indicators, and specific, criteria-based "dictionaries," meant that changes to the criteria or the number of functions could be made easily and quickly by data entry alone. Altering the basic program itself only requires modifying the Paradox code, a task that could be undertaken by anyone familiar with the Paradox programming language.

A third advantage of a computer system is efficiency compared to a paper system. For example, a computer system eliminates the keypunching step in preparing the data for analysis, removing a potential source of error in the data. Data storage becomes a matter of a limited number of computer diskettes in contrast to the square yards typically required for storing paper forms.

Paradox is a menu driven database system. The top level menu viewed by the reviewer lists the various tasks performed by the computer program, including data management and medical record review. The first menu of the review system presents the clinical functions performed by the physicians and reviewed by DRRS. The abstractor is directed to look in the physician's office record, visit by visit, for specific items that relate to a function, such as a diabetic patient or an abnormal blood pressure test. The reviewer then selects the function corresponding to the item from the computer menu and the first of a series of computer screens appears.

The initial screen of each module contains a list of clinical items, and their codes, reviewed by DEMPAQ for the particular function. Once the code of a particular item is entered, the program accesses the appropriate dictionaries with the criteria for the specified clinical item. The next several screens pose questions related to the indicators applicable to that item. For example, for the drug monitoring indicator, the nurse reviewer is asked "Indicate whether each monitoring action was done in the time specified." The monitoring actions required for a particular clinical item appear in a "prompt" box in the center of the screen. Each screen also shows the response options for the question being asked. Reviewer responses are stored in Paradox dictionaries.

The computer system was designed and tested during the spring and summer of 1991. Training sessions at the Maryland PRO (Delmarva Foundation for Medical Care, Inc.) took place in the fall of 1991. These practice sessions generated additional changes to the review system.

The actual review began in November 1991 at DFMC. Initially the first 100 records were reviewed again by a second nurse reviewer and disagreements between the two reviews were discussed among the reviewers. These discussions led to many alterations in the DRRS program.

Sampling

We sampled in two stages: physicians first and then patients within physicians. The samples for each state were drawn separately.

Physician Sample

The physician sampling frame for a state was constructed from National Claims History (formerly Common Working File) claims data for the time period July 1, 1989, through July 30, 1990. A file was prepared of all the physicians whose codes appear on the data set along with data on their specialty, location, and a measure of volume (visits). Once we had a listing of all physicians in the state who submitted at least one claim, we excluded certain categories of physicians from the sampling frame:

- nonprimary care physicians (we retained only those physicians whose most frequently listed specialty was family practice, internal medicine or general practice);
- physicians with fewer than 25 primary Medicare patients per year[1] (we needed to sample at least 25 patients in order to obtain our desired response rate of 20 records per physician); and
- physicians with fewer than 2 claims per quarter (we wished to exclude inactive physicians from the sampling frame on the grounds that inactive physicians neither have enough visits to review nor do they represent mainstream typical care in the state).

After making the necessary exclusions, we stratified our sample by specialty[2]—family practice, general practice, and internal medicine. Stratifying allowed us to oversample specialties that were known to have proportionally fewer members than others. We also implicitly stratified the sample by location by ordering the sampling frame by zip code. This gave a geographic distribution to the sample.

---

[1] Because we lacked a full year's worth of data for one of the carriers (Camp Hill, Pennsylvania), we excluded physicians with fewer than 12 patients per year from this carrier.

[2] For Iowa we created a fourth stratum, osteopathic physicians. Originally we had intended to feedback data to four groups: family physicians; general practice physicians; internists; and osteopathic physicians. During preliminary feedback, we learned that the Iowa physicians preferred to be grouped into three, rather than four, categories.

The actual sampling was conducted by drawing a systematic random sample, with different sampling frequencies for each of the specialties. We assumed a response rate of 33%; it is notoriously difficult to achieve high response rates to physician surveys (Charles et al., 1984; Charles et al, 1985; Peters et al, 1986). We selected a sample of 300 physicians in each state, approximately 100 physicians from each specialty group.

Patient Sample

The sampling frame for the patient sample consisted of all patients for whom the sampled physician was the principle source of primary care. Using the claims data we linked patients and physicians. When two or more primary care physicians provided an identical number of visits to a patient, the amount of ancillary services provided was used to break the tie, thus identifying the physician known as the patient's "primary care source." Overall, approximately 80% of the primary care visits were provided by the patients' primary care source.

We selected 25 patients per physician, assuming that if 80% of the records were found, this would yield 20 charts for review. Patients were identified by drawing a systematic random sample, by physician, with a sampling interval that was proportional to the total number of patients assigned to the physician.

Conducting the Review

Conducting the review involved several steps: inviting the physicians to participate and receiving their records, preparing the chart for review and actually abstracting the record.

Invitation to Physicians

Physicians were invited to participate in the DEMPAQ project through a special letter of invitation mailed by the medical director of each PRO. Prior to the letter of invitation, the PROs in each state had secured agreement of the relevant state medical societies to circulate news articles describing the project and the upcoming review. The letter of invitation contained detailed instructions on how to participate, including the time period for the review and a list of required data. The letter also included assurances that all information would be strictly confidential, i.e., that data linked to the individual physician would never be released. A reimbursement of $5.00 per record was offered to help defray the costs of processing and mailing the records to the PRO. In addition, the letter indicated the PRO's willingness to travel to individual physicians' offices and photocopy the records on-site.

Chart Preparation

Prior to beginning the review, the nurse reviewer was instructed to scan the physician's office record to get a sense of the type of information it contained and to familiarize herself with the facts of the case. While reading through the physician's office record for the first time, the nurse reviewer was instructed to maintain a specially formatted worksheet for recording prevention/screening care done and the dates of abnormal tests. This worksheet facilitated review of the prevention and test functions.

Data Abstraction

Data abstraction took place over a five-month period in each of the three study states. One state used help sheets which listed the decision rules for reviewing commonly found items. For example, the help sheets listed common synonyms for each of the conditions reviewed by DEMPAQ. They also detailed the rules for determining when an item should be reviewed. Nurse reviewers tried to balance difficult (long or hard to read) charts with shorter, legible charts. Comments from the nurse reviewers at the end of the review period identified several strategies for improving the quality of the reviews. These suggestions ranged from scheduling difficult charts for first thing in the morning and reducing the total number of hours allotted to review each day as fatigue becomes an important factor in the late afternoon, to scheduling weekly or twice weekly nurse reviewer discussion groups to cover unusual events or situations (see Chapter 13).

Data Quality Control
Several steps were undertaken to ensure the quality of the data collected. These included a tracking system, a daily file backup system and on-going inter-rater reliability tests.

The tracking system charted the course of the physician's office record through the review process. For each physician, computer generated pages listed the patients sampled and their ID numbers. Spaces after the patient ID number were provided to record the data the chart was received, the date and ID number of the first review(er), the date and ID number of the second review(er) and the date and ID number for the physician peer review(er). Tracking sheets were printed for all sampled physicians; only those for the responding physicians were used. This method of tracking facilitated the identification of nonresponders and of the relevant reviews.

All charts were subject to a primary review. The first 100 charts were subject to a rereview by a second nurse reviewer. The purpose of the rereview was to ensure that all nurse reviewers were interpreting the screens and decision rules in a similar fashion. Discrepancies among nurse reviewers led to discussions of why the discrepancies occurred and thence to substantive revisions to the training manual and sometimes the computer screens. To guarantee compatibility between states, the first 100 cases from Iowa and Alabama were rereviewed by Maryland nurse reviewers. This ensured that Alabama and Iowa reviewers interpreted the criteria in the same way as the Maryland reviewers.

Following the first 100 case rereview, a 5% random sample of remaining cases was subject to a rereview. The purpose of this review was to confirm that nurse reviewers continued to interpret the criteria in a similar fashion. The 5% rereview calibrated review against each state's standard.

The data were collected on nonnetworked PCs. At the end of each day's review session, the data were downloaded to ASCII files and transported to a "master" PC for storage.

Scoring and Rate Calculation
At the completion of the review, each finding for a case is assigned one of five scores: met; acceptable alternative (AA); not applicable; not reviewable; and flag. Met means the record meets the criterion exactly; acceptable alternative indicates the record did not meet the criterion for a set of predefined reasons, e.g., patient refusal; not applicable means the criterion does not apply to this case; not reviewable indicates the time frame for records data was insufficient to review this item for this case; and flag means the case did not meet the criterion, nor any of the categories above. The data are scored by downloading information from the master machine and running a program written in SAS (1988). After scoring, the data are aggregated to create four types of performance rates: clinical item rates; indicator summary rates; function summary rates; and records documentation rates.

Data for the record documentation function are scored differently. The nurse reviewers evaluate the quality of record documentation on a four point scale: 0 = no documentation; 1 = illegible documentation; 2 = minimal documentation; and 3 = adequate documentation.

Clinical Item Rates
Figure 1 on the next page is an example showing how the acceptable performance scores (Met + AA) for each patient are added together and divided by the total number of scores for a doctor (MET + AA + FLAG) to give a performance rate for each criterion (see rows 1 and 2, figure 1). A physician's monitoring rate is calculated by averaging individual patient monitoring scores (see row 3, figure 1).

Figure one
Calculating Rates for a Clinical Item for a Physician
Diabetes Diagnosis

Criteria	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5	Rate*
If on Steroids, review	NA	NA	Met	Flag	Flag	1/3 = 33%
Bl. glu > 250 with symptoms, etc.	NA	NA	Met	Met	Flag	2/3 = 67%
Overall Dx monitor	75%	38%	NA	NA	NA	(75+38)/2 = 57%

* Rate = (Met + AA)/(Met+AA+Flag) where Met=1, AA=1 and Flag=0

The drug and diagnosis functions include monitoring. Each monitoring rate may be derived from up to eight individual monitoring scores. Overall monitoring rates are the percentage of monitoring criteria with met or acceptable alternative findings. In the example below, patient 1 has six met or AA scores on the individual criteria, one flag score and one score which is not included in either the numerator or the denominator. Thus, the overall monitoring rate for patient 1 is 86%.

Figure Two
Calculating Monitoring Rates for a Physician
Diabetes Diagnosis

Criteria	Patient 1	Patient 2	Rate*
Overall Dx Monitor Rate	6/7 = 86%	3/8 = 38%	(86+38)/2=62%
Comment on glucose control	Met	Flag	1/2 = 50%
Wt, pulse,BP	Met	Met	2/2 = 100%
HbA1	Flag	Flag	0/2 = 0%
U/A dipstick for protein	Met	Met	2/2 = 100%
Blood glucose	AA	Met	2/2 = 100%
Diet compliance	AA	Flag	1/2 = 50%
Heart, leg and foot exam	Met	Flag	1/2 = 50%
Fundoscopy	NR	Flag	1/2 = 50%

* For individual criterion; rate = (Met+AA)/(Met + AA + Flag), where Met=1, AA=1 and Flag=0
For overall rate; rate = average of individual criterion rates

Indicator Summary Rates

The indicator summary rates are first calculated for a patient and then averaged across all patients. Figure 3 on the next page depicts how the MET and AA scores for each clinical item are added together and divided by the total number of applicable scores for a patient (MET + AA + FLAG). This gives a patient level performance rate for a particular indicator. The monitoring (drug and diagnosis) indicator rates are calculated by averaging the percentage of monitoring items with met or AA findings across all applicable clinical items.

Figure Three
Diagnosis Indicator Rates
For a Single Patient

Indicator	Diabetes	Hyper-tension	Ischemic Heart Disease	Congestive Heart Failure	COPD	OA	Rate*
Old Dx Monitoring	88	50	33	NA	NA	NA	(88+50+33)/3 = 57%
Evidence for a New Diagnosis	NA	NA	NA	Met	Flag	NA	1/2 = 50%
Search for Treatable Causes	NA	NA	NA	Flag	Flag	NA	0/2 = 0%

* For non-monitoring indicators, rate = (Met + AA)/(Met+AA+Flag) where Met=1, AA=1 and Flag=0
For monitoring indicators, rate = average across all diagnoses

Figure 4 shows how the indicator rates for each patient are averaged to create the physician level indicator rate. The denominator is the number of patients to whom the indicator applies. This method of calculating a summary rate ensures that no single patient will unduly influence the summary rate.

Figure Four
Indicator Summary Rates
For a Physician

Indicator	Patient 1	Patient 2	Patient 3	Rate*
Overall Dx Monitoring Rate	88%	50%	33%	(88 + 50 + 33)/3 = 57%
Evidence for New Diagnosis	67%	100%	NA	(67 + 100)/2 = 84%
Search for Treatable Causes	100%	100%	NA	(100 + 100)/2 = 100%

* Rate = average across all patients

Function Summary Rates

Figure 5 shows how function summary rates are derived. The number of acceptable scores (met + AA) are summed across all clinical items and indicators within a function and divided by the total number of scores for a patient (MET + AA + FLAG). This gives a patient level function performance rate.

Figure Five
Function Summary Rates
For a Single Patient

Indicator	Hypertension	Congestive Heart Failure	COPD	OA	Sum Across Conditions*	Number of Scores
Old Dx Monitoring	.62	NA	NA	.75	1.37	2
Evidence for a New Diagnosis	NA	Flag	Met	NA	1	2
Search for Treatable Causes	NA	Flag	Flag	NA	0	2
Grand Total					2.37	6
Function Score					2.37/6 = 40%	

* Flag = 0, Met and AA = 1

The patient level function summary rates are then averaged (see Figure 6). The denominator is the number of patients to whom the function applies. As with the indicator summary rate, this method of calculating a summary rate ensures that no single patient will unduly influence the summary rate.

Figure six
Diagnosis Function Summary Rates
For a Physician

Patient level diagnosis rates	
Patient 1	68%
Patient 2	50%
Patient 3	54%
Patient 4	37%
Average	52%

Record Documentation
Figure 7 on the next page shows how record documentation scores are calculated for each patient.

Figure Seven
Record Documentation Rates
For a Single Patient

Documentation Component	Visit 1	Visit 2	Visit 8	Visit 9	Rate
Date	3	3	3	3	12/4 = 3.0
Subjective	3	3	2	3	11/4 = 2.75
Objective	3	3	3	3	12/4 = 3.0
Assessment	2	3	3	3	11/4 = 2.75
Plan	3	3	3	3	12/4 = 3.0
Clinician ID	1	1	3	3	8/4 = 2.0

Data are then averaged across patients to create a summary score describing a physician's average level of documentation (see Figure 8). Scores close to zero indicate poor documentation, while scores close to 3 indicate adequate documentation.

Figure Eight
Record Documentation Rates
For a Physician

Documentation Component	Patient 1	Patient 2	Patient 8	Patient 9	Rate
Date	3.0	3.0	3.0	3.0	12/4 = 3.0
Subjective	2.75	1.75	2.5	3.0	10/4 = 2.5
Objective	3.0	1.5	2.25	2.75	9.5/4 = 2.4
Assessment	2.75	1.75	2.75	2.75	10/4 = 2.5
Plan	3.0	2.0	3.0	2.75	10.75/4 = 2.7
Clinician ID	2.0	1.0	1.0	1.0	5/4 = 1.25

Reports

Two types of reports have been developed: one for liaison physicians and one for individual physicians. This chapter describes only the liaison reports. The feedback reports are described as part of the final report for the Feedback and Education Modification of the DEMPAQ Project.

The feedback reports for liaison physicians had four goals:

- Inform advocates for the DEMPAQ project;
- Present preliminary results based on state data;
- Identify strengths and weaknesses of the results; and
- Identify potentially interesting findings that might be used as CME targets.

DEMPAQ fostered working relationships with liaison physicians in each of the three study states. These physicians were kept apprised of the progress of the project at every stage. At the project's outset, many of the liaison physicians were doubtful about the project or even distrustful of us as partners. By the end of the project, most had demonstrated thoughtful, constructive collaboration. By presenting the liaison physicians with preliminary data, we were honoring our commitment to them and providing them with the information they needed to be advocates for the project. The liaison physicians acted as "opinion leaders" for colleagues in their specialties.

A second objective of the reports was to present preliminary results based on the abstraction of physician office records in each state. An example of the liaison feedback tables is shown as Figure 9. The header of the table identifies the function and clinical item being presented. The first column of the table lists the indicators associated with the function. The second column shows the clinical item specific criteria for the indicator. Columns 3-6 show the results of the review. Column 3 shows the number and percent of cases where the criteria were met. Column 4 shows the number and percent of flagged cases. Column 5 shows the number and percent of acceptable alternative cases. For the feedback reports to individual physicians, the met and acceptable alternative data were combined. Column 6 displays the number of cases which were ineligible for the criterion.

The next columns present results of two evaluation studies. The data in Column 7 show the results of the inter-rater reliability. A high percentage of agreement indicates that the information in the record could be consistently abstracted. The validity data were based on structured physician assessments of the flags. The physicians were asked to consider whether there were patient, practice or other factors which explained the flag. In situations where they disagreed with the criteria, they were asked to suggest specific revisions. (For further details of the structured physician review, see Chapter 14). Taken together, the data in these two columns aided the liaison physicians in deciding which of the many DEMPAQ findings were truly interesting.

FIGURE NINE
An example of Liaison Reports for Diagnosis

DIABETES

CODE: 05

Indicator	Criterion	Scored or Raw Data				Inter-rater	Peer Review	Comments
		Met/ Yes (N)	Flag/ No (N)	AA/ Other (N)	Ineligible			
Old Diagnosis Monitoring	Denominator	102	102	102	111	28	18	
	Overall Score:	101	101	101				
		28% (28)	72% (73)	0% (0)	10		Flag justified = 10 Patient factors = 2 Extenuating circumstances = 1 Practice factors = 1 Manual revision = 1 Criteria revision = 3	
		100	100	100				
	Comment on blood glucose control q 6 mos	27% (27)	56% (56)	17% (17)	11	82%		
		101	101	101				
	Weight, pulse, BP - q 6 mos	18% (18)	63% (64)	19% (19)	10	86%	**1a-2a-3a. Pulse should be deleted.**	
		101	101	101				
	HgbA1C or fructosamine - q 6 mos	5% (5)	77% (78)	18% (18)	10	89%	**2b. Only when symptoms show patient is less controlled should HgbA1C be done q 6 mos.**	

NR = Not Reviewed
NA = Not Applicable

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

DIAGNOSIS ANALYSIS - DIABETES - IOWA - NOVEMBER 3, 1993

CODE: 05

Indicator	Criterion	Scored or Raw Data			Ineli-gible	Inter-rater	Peer Review	Comments
		Met/ Yes (N)	Flag/ No (N)	AA/ Other (N)				
		101	101	101				
	Urinalysis dipstick for protein and microscopic exam of sediment - q 6 mos	2% (2)	82% (83)	16% (16)	10	93%	*1. Should be acceptable to do within 6 mos.* **1b. U/A should be done once/yr.** **2c-3b. Only when symptoms show patient is less controlled should U/A be done q 6 mos.**	Note: Alabama reviewers interpreted criteria as within 6 mos.
		100	100	100				
	Blood glucose - q 6 mos	40% (40)	41% (41)	19% (19)	11	89%		
		100	100	100				
	Document dietary compliance - once	37% (37)	49% (49)	14% (14)	11	89%		
		100	100	100				
	Heart exam, leg and foot exam - once	43% (43)	42% (42)	15% (15)	11	86%		
		100	100	100				
	Funduscopy - once	17% (17)	60% (60)	23% (23)	11	93%	**2d. Not useful in primary care because cannot see enough to say if problem.**	

NR = Not Reviewed
NA = Not Applicable

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

DIAGNOSIS ANALYSIS - DIABETES - IOWA - NOVEMBER 3, 1993

CODE: 05

Indicator	Criterion	Scored or Raw Data				Inter-rater	Peer Review	Comments
		Met/ Yes (N)	Flag/ No (N)	AA/ Other (N)	Ineli-gible			
Evidence for new diagnosis	Denominator	9	9	9	111	2	17	
	Blood glucose >250 with symptoms of DM OR two FBS >140 OR blood glucose levels during GTT >250 at 2 hrs and some point between 0 and 2 hrs	44% (4)	56% (5)	0% (0)	102	100%	Flag justified = 4 Manual revision = 4 Extenuating circumstances = 1 Practice factors = 1 Incomplete rule = 1 Criteria revision = 6 1. GTT not reported, but dx and therapy indicate positive. *2-3. On a new diagnosis, a single fasting BS >200 should be acceptable. Two random BS >200 is acceptable; may not have weight loss.* **4. Rigid rule to document associated symptoms should not be necessary to establish diagnosis.** **5. Drop GTT requirement.** **6. Mirror ADA.**	
Search for Treatable Causes	Denominator	9	9	9	111	1	0	
	If on steroids, possible potentiating role reviewed.	33% (3)	67% (6)	0% (0)	102	100%		

NR = Not Reviewed
NA = Not Applicable

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (**bold**), and *Maryland* (*italics*).

Improving the DRRS: Suggestions for the Future

The process of conducting the review uncovered numerous areas where the record review technology could be improved. Some of these areas are discussed in other chapters (see Chapters 14 and 15 for summaries of peer review and liaison physician suggestions). The table below summarizes changes that need to be undertaken should the DRRS be used again. The suggestions are presented in order of importance.

Table 1

Prioritized Listing of Proposed DRRS System Changes

DRRS Component	Proposed Change
Criteria	1. Revise criteria based on recommendations in chapters 14 and 15. This is probably the most important step in the revision process.
Decision Rules	1. Revise decision rules based on recommendations in chapters 13 and 14. This includes revising the monitoring time frames to either: • within the specified time period; or • a certain number of times at a specified number of months (weeks) apart.
DRRS Computer Screens	1. Revise based on recommendations in chapter 13
Scoring	1. Include scoring options for the following: • patient did not keep scheduled appointment • physician advised patient to receive (test, x-ray, etc.), not known if patient did
Test Function	1. Eliminate test indications indicator. There was very little variability on this item. The overwhelming majority of tests were appropriately ordered. 2. Revise the appropriate response screen. Instead of asking the reviewer to examine the record for a sequence of actions, list the actions and provide a field for the reviewer to note the presence or absence of each action. The scoring program would then calculate the score either on the basis of: • the percentage of actions performed; or • which actions are performed. 3. Eliminate test urgent response indicator. The number of cases was infinitesimal.

DRRS Component	Proposed Change
Diagnosis Function	1. Revise the treatable causes screen. Provide a field for the reviewer to note the presence or absence of each action. The scoring program would then calculate the score on the basis of: • the percentage of searches made; or • which searches were made. 2. Consider adding a new indicator - "Appropriate workup of a new diagnosis"
Drugs	1. Drop contingent monitoring. The sample size was consistently too small. 2. Consider dropping the two contraindications indicators. The performance was quite good on these two indicators with little variability. 3. If the contraindications indicators are retained, provide a field for each contraindication and ask the reviewer to note which contraindication was present.
Procedures	1. Drop. Too few are done in the physician's office.
Complaint Function	1. Re-instate the complaint function. This potentially important function was dropped from the pilot because the review took too long. The indicator would be "Appropriate workup of a complaint". The criteria would be divided into two sets: new complaint, pre-existing complaint with exacerbation. The response screens would contain lists of actions such as tests ordered, physical exam findings, or treatments ordered. The reviewer would check off which actions were done by the physician.

Chapter 12
Comparison of Participants and Nonparticipants in the DEMPAQ Record Review

Abstract

Methods
We compared the characteristics of physicians in each study state who agreed to submit records for the record review with the characteristics of physicians who declined to participate. Analyses were performed at the bivariate level using chi square or t-test tests of significance as appropriate. Logical regression was performed using participation status (yes/no) as the dependent variable.

Results
There were no differences between responders and nonresponders in terms of age, type of practice (solo versus group), geographic location (urban versus rural), quality of care (as measured by the claims), and hospital use.

There were differences in participation by state (Iowa physicians were less likely to participate), specialty (general practitioners were less likely to participate than family practitioners or internists) and average total resource units per patient (physicians with costlier case loads were more likely to participate).

Conclusion
These findings suggest that our results may be generalized with regard to age, solo and group practice, urban and rural settings, Medicare practice size, quality of care and total hospital expenditures.

Introduction/Statement of the Problem

One of the two review methods developed for DEMPAQ depended on primary care physicians volunteering to submit copies of specified random samples of their Medicare patients. It is important to understand what differences exist between participating and nonparticipating physicians in order to understand the extent to which results from that method can be generalized.

Methods

Data files were prepared at Johns Hopkins University which contained a number of physician characteristics. (These data files had been used for other analyses found in Chapters 6, 7 and 8 of this final report.) We extracted the files that pertained to primary care physicians who had been invited to participate in the medical record review portion.

Because the study had been conducted in a time frame when UPINs were not universally available, physician identity was confounded by the facts that multiple physicians could use the same billing number and single physicians could use multiple billing numbers (see the discussion in Chapters 4 and 5). For the purposes of the overall study we assigned physicians an identifier as a primary care source (PCS). The analysis below is based on the PCS identifier.

There were 850 PCSs on the original provider file. Nine PCSs were dropped because each had conflicting data as to whether they participated or not, leaving 841. Seven hundred eleven of these had the necessary additional information (analysis variables) for this study. One PCS was removed because the same number appeared in both the Alabama and Maryland files. Of the 710 PCSs, 704 had unique specialty codes on the provider file. The six PCSs with conflicting specialty information were dropped from the analysis.

Variables available for the analysis included:

- State (Alabama, Iowa, Maryland);
- Specialty (general practice, family practice, internal medicine);
- Solo vs. group practice;
- Physician age (for Alabama and Maryland only);
- Metro vs. nonmetro practice location (based on SMSA assignment of practice zip code);
- Practice size (divided in tertiles based on distribution of number of Medicare patients identified in this study);
- Quality of care (a continuous variable ranging from 0 to 1.0, reflecting the percent of patients with a given condition receiving recommended care for that condition);
- Average total hospital resource units (average of total resource units [standardized charges] incurred by patients assigned to that PCS); and
- Average total resource units/patient (average of total resource units [standardized charges] incurred by patients assigned to that PCS);

Analyses were performed at the bivariate level using chi square or t-test tests of significance as appropriate. Logistical regression was performed including the significant variables found in the bivariate analyses. All analyses were performed for the overall sample and within each of the three states.

Results

Over one quarter (27.3%) of the physicians who were invited to participate, submitted medical records for review. The participation rate varied by state as shown below, with Iowa having a lower participation rate (18.9%) than Alabama (31.4%) or Maryland (31.5%).

Figure 1

Participation Rate by State

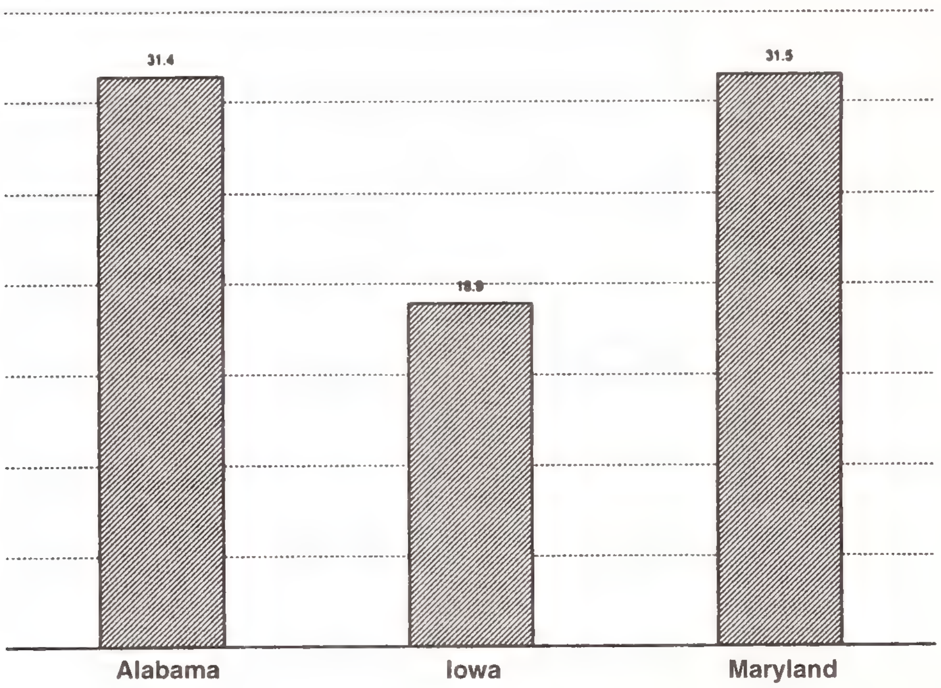


Table 1
Participation Rate by State

	AL	IA	MD	
Non-Participants	68.7% (157)	81.1% (189)	68.6% (170)	72.7% (516)
Participants	31.4% (72)	18.9% (44)	31.5% (78)	27.3% (194)
	229	233	248	710

Figures 2 and 3 which follow show the variation in participation by specialty. General practitioners were less likely to participate than family practitioners or internists. This pattern remained true in each of the three states.

Figure 2

Participation Rate by Specialty - Overall

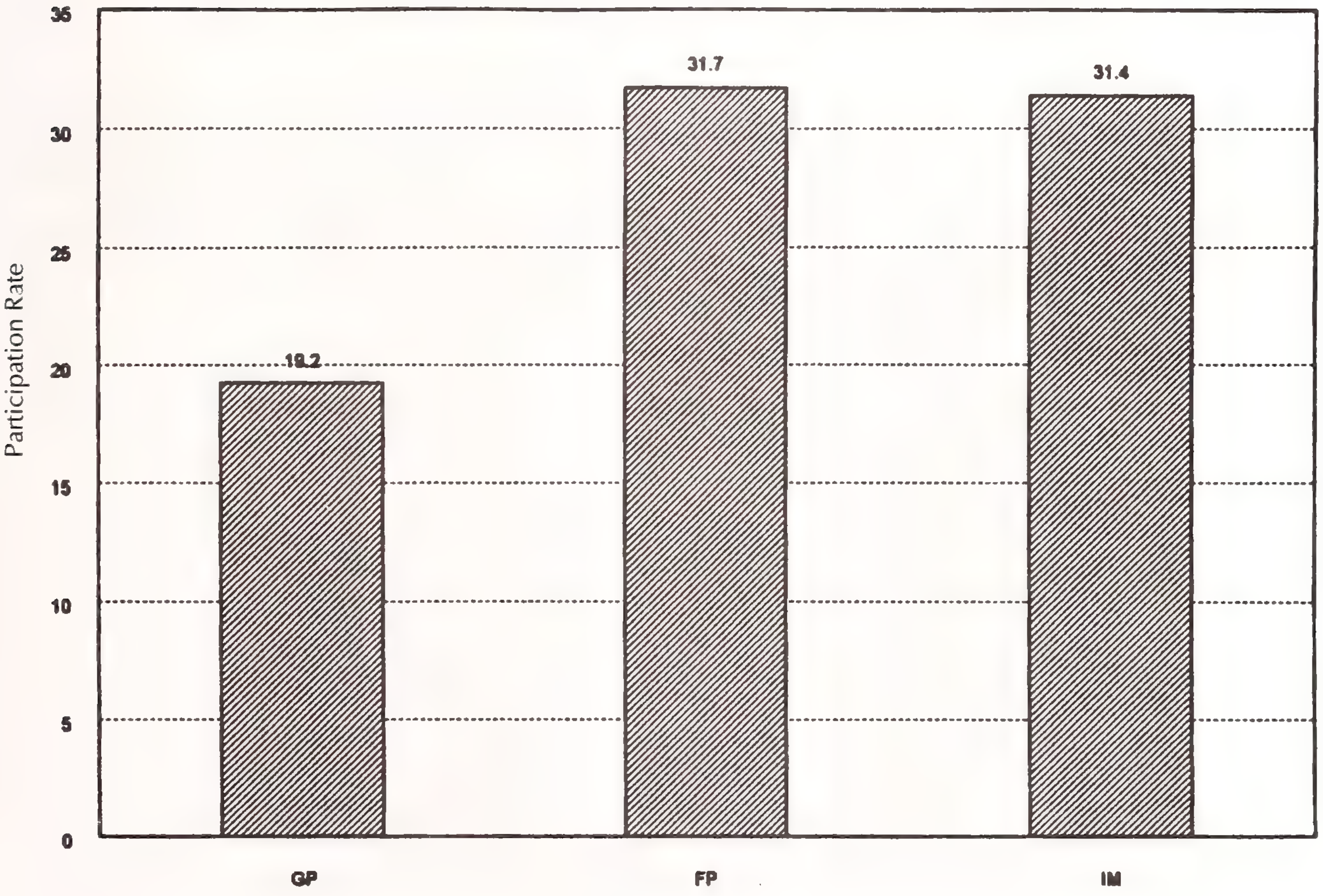
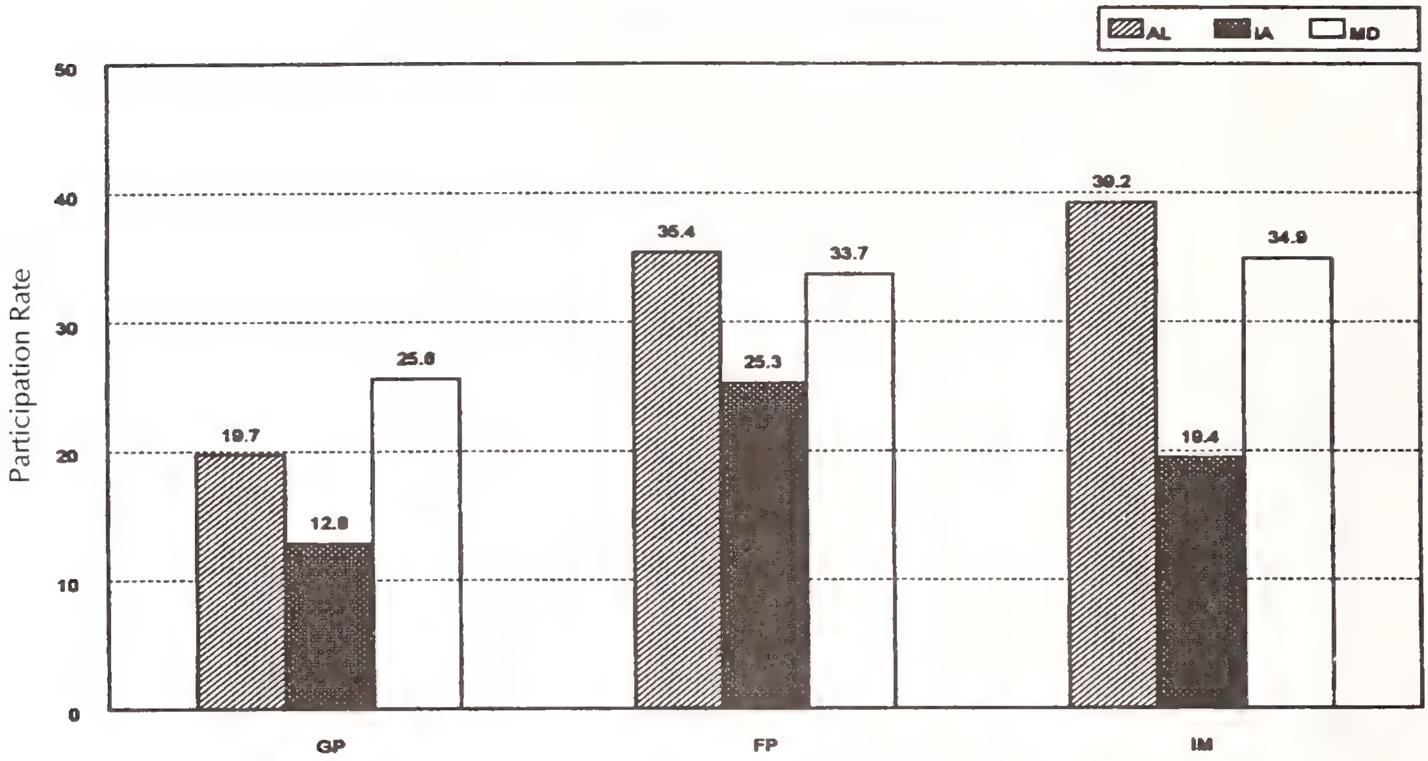


Figure 3

Participation Rate by Specialty Within Each State



There was no significant difference in participation rates between physicians in solo practice and physicians in group practice (either single specialty or multispecialty). However, one should be cautious in interpreting these results since the PCS trimming rules described in Chapter 5 probably lead to greater trimming of group practices than solo practices. The overall distribution of solo practice (83%) is not the same as that found in the full study sample, in which the distribution was 74% for Alabama, 62% for Iowa and 77% for Maryland (information provided by Johns Hopkins University).

Table 2
Participation Rate by Type of Practice (Group or Solo)

	Group	Solo	
Non-Participants	67.2% (82)	73.8% (434)	72.7% (516)
Participants	32.8% (40)	26.2% (154)	27.3% (194)
	122	588	710
	17.2%	82.8%	710

As shown in Table 3, there was no significant difference by physician's age in either of the two states for which we had this information.

Table 3
Participation by Age

	Mean Age (SD)	Significance
Alabama		
Non-participant	48.2 (12.8)	
Participant	45.8 (12.5)	ns
Maryland		
Non-participant	53.8 (12.4)	
Participant	50.1 (12.7)	ns

Figure 4 shows the participation rate by metropolitan compared to nonmetropolitan SMSAs. There was no difference in participation rates by type of metropolitan area.

Figure 4

Participation Rate by Metropolitan Type

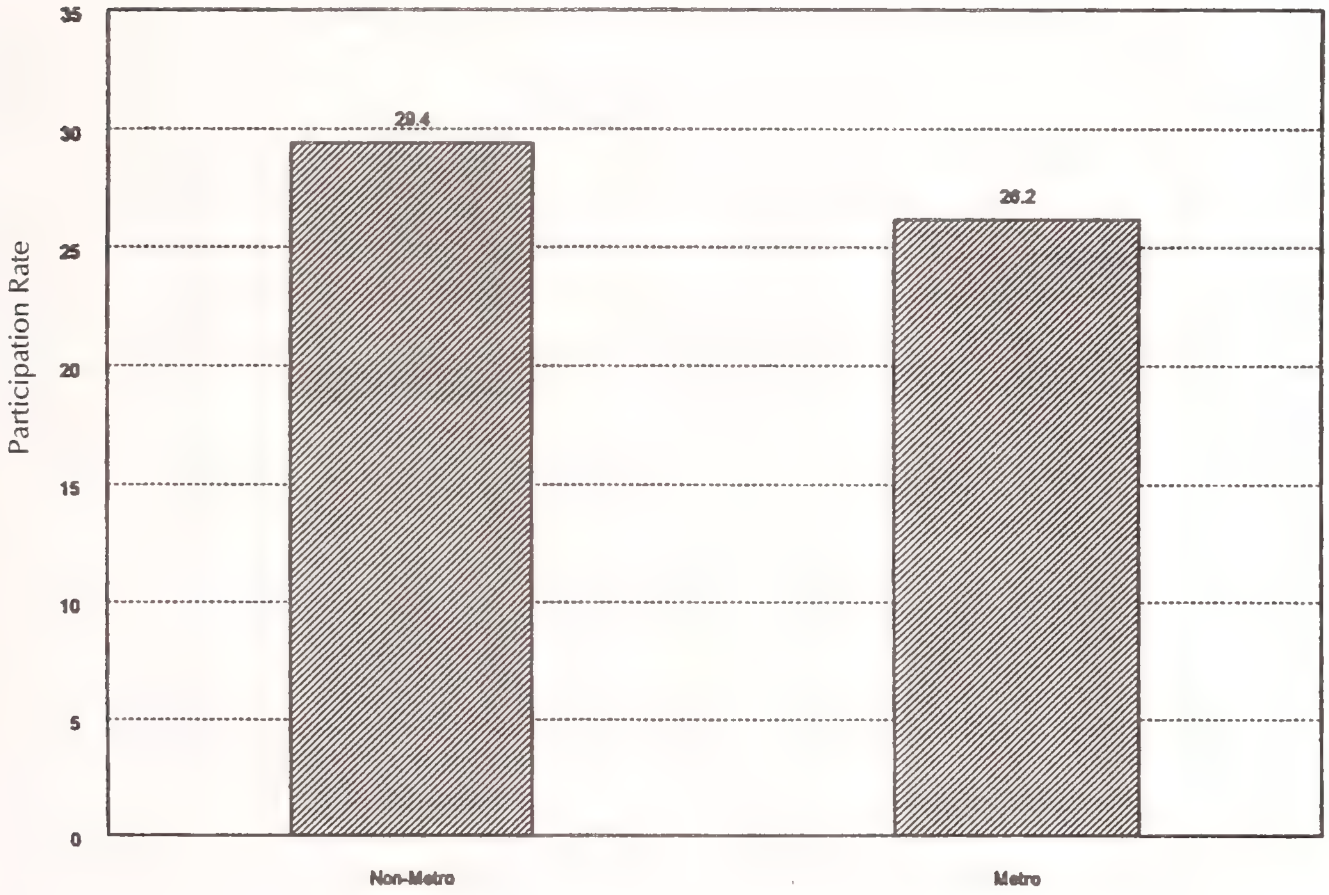


Table 4 shows the variation in participation by quality of care. Although the measure is a crude one, calculated by averaging the sum of recommended care items for six different conditions, there appears to be no significant difference between participants and nonparticipants on this measure.

Table 4
Participation by Quality of Care

	Mean quality score (SD)	Significance
Overall		
Non-participant	0.45(.36)	
Participant	0.46(.36)	ns
Alabama		
Non-participant	0.46(.36)	
Participant	0.47(.35)	ns
Iowa		
Non-participant	0.42(.36)	
Participant	0.42(.35)	ns
Maryland		
Non-participant	0.48(.37)	
Participant	0.47(.36)	ns

Table 5 shows the variation in participation by mean total hospital resource units per patient. This measure was created by substituting average state charges for each hospital procedure for the actual charges. There was no significant difference on this measure.

Table 5
Participation by Mean Total Hospital Resource Units per Patient

	Mean Hospital Resource Units (SD)	Significance
Overall		
Non-participant	864,433 (1,383,255)	
Participant	995,363 (1,398,214)	ns
Alabama		
Non-participant	1,671,180(2,234,673)	
Participant	1,823,657(1,992,570)	ns
Iowa		
Non-participant	626,485 (445,476)	
Participant	654,249 (418,345)	ns
Maryland		
Non-participant	374,813 (302,172)	
Participant	423,207 (319,198)	ns

Table 6 shows the variation in participation by mean total resource units per patient. Overall, participants have higher mean total resource units than nonparticipants, e.g. participating physicians have higher standardized charges per patient than nonparticipating physicians. Note that this measure does not control for patient case-mix.

Table 6
Participation by Mean Total Resource Units per Patient

	Mean Total Resource Units/Patient (SD)	Significance
Overall		
Non-participant	5,323 (2,659)	
Participant	5,636 (3,170)	.002
Alabama		
Non-participant	7,009 (3,097)	
Participant	7,338 (4,034)	
Iowa		
Non-participant	4,853 (1,963)	
Participant	4,838 (2,174)	
Maryland		
Non-participant	4,290 (2,109)	
Participant	4,516 (1,791)	

We also performed logistic regressions for the overall sample and separately for each state. We examined all variables and the interaction terms for specialty times the other variables. None of the interaction terms was significant. In the overall sample only specialty and state remained significant. For Alabama, both specialty and SMSA were significant, with nonmetro physicians more likely to participate (38.8% vs. 26.9%, p-val. = .01).

Since none of the continuous variables or interaction terms was significant in the logistic regressions, we summarize the significant results in cross-tab tables. General Practice physicians were significantly less likely to participate than Family Practice or Internal Medicine physicians, but the rates for Family Practice and Internal Medicine physicians were not significantly different from each other.

Table 7

Logistic Regression Results -- Overall Sample

Variable	Logistic regression coefficient	*p*	Adjusted odds ratio
General Practice (dummy)			
Family Practice	-0.64	.004	0.53
Internal Medicine	-0.62	.005	0.54
Alabama (dummy)			
Iowa	0.66	.003	1.94
Maryland	0.00	.983	1.00

Table 8

Logistic Regression Results -- Alabama

Variable	Logistic regression coefficient	*p*	Adjusted odds ratio
General Practice (dummy)			
Family Practice	-0.87	0.025	0.42
Internal Medicine	-1.16	0.005	0.32
SMSA	0.77	0.016	1.16

Table 9

Logistic Regression Results -- Iowa

Variable	Logistic regression coefficient	*p*	Adjusted odds ratio
General Practice (dummy)			
Family Practice	-0.84	0.045	0.43
Internal Medicine	-0.50	0.257	0.61

Table 10

Logistic Regression Results -- Maryland

Variable	Linear logistic regression coefficient	*p*	Adjusted odds ratio
General Practice (dummy)			
Family Practice	-0.35	0.308	0.70
Internal Medicine	-0.46	0.184	0.63

Conclusions

In summary, the bivariate analysis shows only two significant variables: state (Iowa had a lower participation rate than Alabama or Maryland), and specialty (General Practice physicians are less likely to participate).

Many structural and demographic variables that we had expected to play a role were not significant. These findings suggest that our results may be generalizable with regard to age, solo and group practice, urban and rural settings, Medicare practice size, quality of care and total hospital expenditures.

Of course the model is limited because there could be other important variables that we could not measure. These characteristics, such as extent of medical leadership or other measures of quality, could still be important predictors of participation.

Chapter 13
Evaluating DRRS: Inter-rater Reliability for Data Abstraction

Abstract
The reliability of a data abstraction instrument contributes to its ultimate validity. A tool that produces the same results each time it is used to review a particular chart is more believable than one that produces different results.

Methods
Nurse reviewers in each state conducted a rereview of the first 100 charts subjected to the DEMPAQ review process. After the initial review, a 5% sample of charts (one randomly selected per physician) was reviewed. In addition, at the end of the review period, comments and suggestions were solicited from the nurse reviewers about how to improve the review.

Results
Data from the first 100 cases reviewed in each state led to changes and modifications of the abstracting instructions, the computer screens, and the dictionary presentation. The ongoing 5% review confirmed that reviewers continued to interpret the review instructions consistently.

Discussion
Reviewers became committed to DRRS, a fact which probably contributed to the high inter-rater reliability for most aspects of the system. Not only did the reviewers become involved in the "stories" of the patients, they felt they were contributing to development of a useful review tool. As a result of the involvement reviewers offered many excellent suggestions for improving the review instrument (Appendix 4).

This chapter presents the results of a reliability evaluation of the record review instrument. It also includes a summary of the specific suggestions for revising the DRRS to improve inter-rater reliability that were made by the nurse reviewers in each state.

Reliability speaks to the reproducibility of a result (Abramson, 1990). A reliable measurement is one that produces the same results regardless of time, place or person conducting the measurement. We undertook an assessment of the reliability of the DRRS for several reasons. First, we recognized that as a new, untested instrument, DRRS probably had places where the screens were not clear, the instructions were not clear, etc. These problems could very easily lead to different interpretations by reviewers and would need to be cleared up if the review were to proceed. Second, in order for physicians to believe the results of DRRS, we had to demonstrate that the system could achieve the same result with different reviewers. Finally, since the review was taking place in three different states, we needed to ensure that reviewers in each state were interpreting the criteria in the same way.

Methods

This section covers the sampling of cases for review, using the initial results of the review to iron out "bugs," and calculating the kappa scores for agreement.

Sampling—First 100. The goal of the rereview of the first 100 cases was to provide a shakedown period for the reviewers. As the nurse reviewers "went live" with the review, we expected to find numerous problems with the instruction and instrument that had not been uncovered before. Because of this training feature, we anticipated that the results of the first 100 case review would not be not be used in the final data analysis.

To minimize the impact on any one physician's scores of dropping cases, we randomly sampled two patients per physician for the first 100 case review. The list of two random patients per physician was then given to the review coordinator in each state with instructions to use the first case listed for a physician unless the physician had not submitted a record for that patient. Then the coordinator selected the physician's second listed patient.

Nurse reviewers from Maryland conducted the rereview of the Maryland cases, as well as the first 100 cases in Alabama and Iowa. This was to ensure that reviewers in the latter two states interpreted the criteria the same way as reviewers in Maryland. The actual final number of cases reviewed for the "first 100" equaled the number of participating physicians in each state: 91 in Maryland; 92 in Alabama; and 53 in Iowa.

Sampling—Ongoing 5%. As with the first 100 review, the ongoing 5% review sampled two cases per physician. The review coordinator always chose the first case for review, unless that record had not been submitted by the participating physician. In that circumstance, the second case was selected for review.

Feedback to Reviewers. During the review process, results of inter-rater comparisons were fed back to reviewers in each state with the aim of continuously improving the reliability of the review. The initial reports were the most crucial as they aided in the identification of areas where the manual, computer instruction, or training was not clear. These problems needed to be rectified before the main review could proceed.

Inter-rater comparisons were made with a SAS program using PROC COMPARE which produced data and tables pinpointing the areas where the reviewers scored items differently. The presence of a scoring difference meant that reviewers did not agree on how to evaluate aspects of a particular case.

The data were summarized in a report showing the percent agreement on each item scored. Accompanying the report were pages of detailed SAS printout. The detail provided by the report showed the review coordinator and the reviewers exactly where the disagreement occurred and facilitated discussion about the specific reasons for the disagreement and how the problems could be fixed.

Calculation of Kappas. Kappa statistics were calculated to show the degree of inter-rater reliability of the abstraction process. Statistics were calculated on scored data since this is the form in which information is fed back to physicians. For purposes of this analysis the scores of "met" and "acceptable alternative" were collapsed into the met category to mirror the manner in which the data were reported back to physicians. Not applicable scores (NA) and not reviewable scores (NR) were excluded from the analysis. This produced a binary "score" variable for the kappa analysis, with two values—met and flag.

Kappas for drug and diagnosis monitoring were calculated somewhat differently. Monitoring scores were continuous, ranging from 0 to 100%. For purposes of reporting back to physicians, the monitoring scores were divided into quartiles. For purposes of kappa analysis, the scores were divided into four groups, e.g., scores between 0 and 24 were placed in group 1 and scores between 25 and 49 were placed in group 2, etc.

Kappas have been calculated only for the ongoing 5% data versus the main study data. The first 100 review was part of the "shake down" period and does not truly reflect the inter-rater agreement of the instrument.

Results

Tables 1 through 9 summarize the results of the inter-rater reliability analysis between the ongoing 5% rereview and the main study, by state and function. Each table displays the number of cases rereviewed during the ongoing 5% review, the percent agreement between reviewers and the kappa statistic for agreement. Fleiss (1981) reports the following guidelines for interpreting kappa values:

> .75 excellent agreement beyond chance
.4–.75 good agreement beyond chance
< .4 poor agreement beyond chance

When interpreting the kappa values, the reader needs to keep in mind two situations that produce "unexpected" kappa values.

1. Calculation of kappa statistics requires data in more than one cell in a two by two table. If all the data are in a single cell,[1] the percent agreement will be 100% and a kappa statistic will be missing.
2. If data are concentrated in a single cell,[2] the percent agreement may be high and the kappa statistic low.

Diagnosis

For diagnosis the inter-rater reliability was generally excellent. The percent agreement was consistently above 90% for locating a diagnosis and finding evidence to support a new diagnosis. Kappas show that agreement was excellent in Maryland and Alabama and good in Iowa. In general, reviewers in Iowa had more difficulty locating all the diagnoses than reviewers in the other two states. This may reflect regional differences in documenting in the record.

For monitoring, the median percent agreement was 72%. The kappa statistics for overall agreement on monitoring indicates good agreement in all three states (.43–.73). Finding monitoring actions for some diagnoses seemed to present problems. For example, COPD monitoring was problematic in all three states (kappa 0–.5), however, the numbers are small (3 to 9 cases).

---

[1] The figure below illustrates situations where kappa cannot be calculated.

		New Drug Indications, Rater #2	
		Found	Not Found
New Drug Indications, Rater #1	Found	5	0
	Not Found	0	0

[2] The figure below illustrates situations where the percent agreement will be high and kappa will be low.

		Prevention, Rater #2	
		Met	Flag
Prevention, Rater #1	Met	106	5
	Flag	2	2

% Agreement = 94%
Kappa = .33

Table One
Three state inter-rater reliability for locating a diagnosis

Diagnosis	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Hypertension	132	96%	.91	115	98%	.96	120	91%	.79
Ischemic Heart Disease	132	100%	1.0	115	100%	1.0	120	94%	.56
Osteoarthritis	132	94%	.57	115	97%	.91	120	88%	.59
COPD	132	98%	.91	115	97%	.81	120	90%	.59
Diabetes	132	98%	.92	115	100%	1.0	120	94%	.71
Congestive Heart Failure	132	99%	.95	115	99%	.94	120	95%	.71
Overall	132	93%	.91	115	96%	.94	120	77%	.70

Table Two
Three state inter-rater reliability for new diagnosis evidence

Diagnosis	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Hypertension	1	100%	-	1	100%	-	2	100%	1.0
Ischemic Heart Disease	5	100%	1.0	0	-	-	1	100%	-
Osteoarthritis	4	100%	1.0	4	100%	-	6	83%	.67
COPD	4	100%	1.0	3	100%	-	0	-	-
Diabetes	3	67%	0.0	1	100%	-	1	100%	-
Congestive Heart Failure	5	100%	1.0	2	100%	-	2	100%	-
Overall	22	95%	.83	11	100%	1.0	12	75%	.53

Table Three
Three state inter-rater reliability for diagnosis monitoring

Diagnosis	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Hypertension	37	65%	.42	40	85%	.74	25	48%	.23
Ischemic Heart Disease	3	100%	1.0	9	100%	-	4	50%	.2
Osteoarthritis	25	72%	.54	24	79%	.22	15	73%	.38
COPD	5	60%	0.0	3	67%	0.0	9	78%	.5
Diabetes	10	50%	.32	13	77%	.67	9	67%	.54
Congestive Heart Failure	6	100%		6	100%	-	4	100%	1.0
Overall	86	67%	.52	95	84%	.73	66	64%	.43

Tests

Percent agreement on locating an abnormal test was consistently high for selected tests, while the overall percent agreement averaged about 75%. Kappas for agreement on test finding were at the high end of "good" agreement, with Alabama showing overall "excellent" agreement. Reviewers seemed to have more difficulty determining an abnormal ECG (kappa .46–.63) than other tests. This may reflect problems determining whether the ECG abnormality was new or preexisting (the criteria limited review to new abnormalities).

Inter-rater reliability for response to an abnormal test was excellent. This finding was unexpected given the complexity of the sequences of actions the nurse reviewers had to search for.

Table Four
Three state inter-rater reliability
Locating a test - Selected tests

Test	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Blood pressure	114	90%	.74	86	93%	.79	114	88%	.65
Cholesterol	114	89%	.72	86	93%	.74	114	92%	.67
Blood glucose	114	96%	.68	86	99%	.92	114	94%	.63
UA Dipstick	114	92%	.60	86	95%	.81	114	97%	.86
ECG	114	90%	.63	86	93%	.46	114	91%	.53
Overall	114	75%	.70	86	78%	.75	114	67%	.62

Table Five
Three state inter-rater reliability
Response to an Abnormal Test - Selected Tests

Test	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Blood pressure	8	75%	.5	5	80%	.55	5	100%	1.0
Cholesterol	16	94%	.76	11	100%	1.0	6	67%	-.2
Blood glucose	5	100%	1.0	6	100%	1.0	6	100%	1.0
Creatinine	0	-	-	4	100%	1.0	3	100%	-
UA Dipstick	8	88%	.75	8	88%	.75	9	89%	.78
Overall	46	89%	.78	37	92%	.83	32	91%	.8

Drugs

Inter-rater reliability for locating selected drugs was excellent in Alabama (kappa range .85–1.0) and good in Maryland (range .53–.75) and Iowa (range .64–.85). Percent agreement on new drug indications was generally high. Kappas could not be calculated for many of the drugs because of small sample sizes and lack of rating categories (see the introduction to this section for explanation of these issues). Drug monitoring appeared to have slightly lower reliability, although the agreement was good (overall kappas .48–.68). The number of cases reviewed per drug was typically small, a factor which may have influenced the scores.

Table Six
Three State Inter-rater Reliability for Locating a Drug

Drug	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Ace Inhibitors	292	97%	.67	259	100%	1.0	297	96%	.54
Beta blockers	292	96%	.68	259	99%	.91	297	97%	.65
Calcium Channel Blockers	292	95%	.53	259	99%	.9	297	97%	.75
Loop Diuretic	292	96%	.71	259	100%	1.0	297	97%	.67
Insulin	292	99%	.57	259	99%	.93	297	99%	.8
Oral Hypoglycemic	292	99%	.74	259	100%	1.0	297	99%	.85
NSAID	292	94%	.71	259	97%	.87	297	93%	.64
Inhaled Beta Adrenergic	292	99%	.66	259	100%	1.0	297	99%	.57
Cephalosporin	292	97%	.75	259	99%	.97	297	94%	.71
Sulfonamide + Trimethroprim	292	98%	.73	259	99%	.85	297	97%	.70

Table Seven
Three State Inter-rater Reliability for New Drug Indications

Drug	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Ace Inhibitors	2	100%	-	6	100%	-	0	-	-
Beta blockers	1	100%	-	0	-	-	0	-	-
Calcium Channel Blockers	1	100%	-	5	100%	-	2	100%	-
Loop Diuretic	9	89%	.61	3	100%	-	5	80%	0.0
Insulin	0	-	-	0	-	-	0	-	-
Oral Hypoglycemic	2	100%	1.0	1	100%	-	0	-	-
NSAID	13	92%	0.0	24	88%	0.0	11	45%	0.0
Inhaled Beta Adrenergic	2	100%	1.0	0	-	-	1	100%	-
Cephalosporin	10	60%	-.18	17	94%	0.0	18	61%	.09
Sulfonamide + Trimethroprim	6	83%	0.0	7	100%	-	8	88%	.71
Overall	94	88%	.72	110	92%	.82	78	72%	.44

Table Eight
Three state Inter-rater Reliability for Drug Monitoring

Drug	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Ace Inhibitors	8	88%	.8	17	82%	.76	2	50%	0.0
Beta blockers	12	58%	.33	5	80%	.72	7	71%	0.0
Calcium Channel Blockers	9	78%	.57	11	72%	.56	11	73%	.57
Loop Diuretic	12	83%	.6	7	86%	.77	9	100%	1.0
Insulin	2	50%	.33	6	33%	.2	2	0%	0.0
Oral Hypoglycemic	5	60%	.23	4	25%	-.2	7	71%	.22
NSAID	13	85%	.75	12	92%	.87	11	36%	.09
Inhaled Beta Adrenergic	3	100%	-	1	100%	-	2	100%	-
Cephalosporin	0	-	-	0	-	-	0	-	-
Sulfonamide + Trimethroprim	0	-	-	0	-	-	0	-	-
Overall	107	76%	.62	116	78%	.68	88	65%	.48

Prevention

Prevention agreement was generally excellent, although the kappas for blood pressure were rather low in Maryland and Iowa. This stems from the distribution of scores in the 2 x 2 table, rather than from a general lack of agreement (see introduction to this section for explanation). Similar distribution effects may be found in the follow-up data.

Complete data for all three states are shown in Appendices 1 through 3.

Appendix 4 shows the specific suggestions made by the reviewers in each state to improve the inter-rater reliability.

Table Nine

Three state inter-rater reliability
Preventive Services

	Maryland			Alabama			Iowa		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Blood pressure	106	97%	.39	92	100%	1.0	115	94%	.33
Influenza Immunization	106	92%	.82	90	93%	.85	111	96%	.90
Mammogram	25	96%	.83	36	100%	1.0	40	88%	.72
Breast Exam	64	95%	.84	59	98%	.96	67	96%	.89

Discussion

The reviewers found the DRRS review engaging and stimulating. Not only did the reviewers become involved in the "stories" of the patients, they felt they were contributing to the development of a useful review tool. These factors probably contributed to the generally excellent inter-rater reliability. Inter-rater reliability for monitoring, although "good," was somewhat lower than the reliability of the other aspects of the system. This may reflect the fact that monitoring scores are a composite of up to eight individual monitoring actions. Perhaps a more accurate way of capturing the reliability of monitoring would be to calculate kappas for each action. This would have the added benefit of aiding in the detection of items which cause difficulties for the abstractors.

The first 100 case review provided essential guidance to the reviewers about areas where abstraction was problematic. The detailed reports prepared at this point showed the review coordinator and the nurse reviewers exactly where disagreement occurred and facilitated discussion about the specific reasons for the disagreement and how the problems could be fixed.

Appendix 1

Diagnoses - Maryland

Diagnosis	Agreement on Diagnosis			If agree on Dx, then agreement on Status			If agree on status then agree diagnosis monitoring			If agree on status then agree on new diagnosis indications			If agree on status then agree on search for treatable causes		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Hypertension	132	96%	.91	41	93%	.36	37	65%	.42	1	100%	1.0	0		
Ischemic Heart Disease	132	100%	1.0	10	80%	.6	3	100%	1.0	5	100%	1.0	4	100%	1.0
Osteoarthritis	132	94%	.57	37	78%	.36	25	72%	.54	4	100%	1.0	4	100%	1.0
COPD	132	98%	.91	11	50%	.63	5	60%	0.0	4	100%	1.0	3	100%	1.0
Diabetes	132	98%	.92	13	100%	1.0	10	50%	.32	3	67%	0.0	3	100%	1.0
Congestive Heart Failure	132	99%	.95	11	100%	1.0	6	100%		5	100%	1.0	4	100%	1.0
Overall	132	93%	.91	123	88%	.67	86	67%	.52	22	95%	.83	18	100%	1.0

Tests - Maryland

Test	Agreement on Test			If agree on test, then agreement on test indications			If agree on test then agreement on test urgent response			If agree on test then agreement on appropriate follow-up		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Blood pressure	114	90%	.74	23	100%	-	0	-	-	8	75%	.5
Hgb/Hct	114	100%	1.0	2	100%	-	0	-	-	2	100%	1.0
Cholesterol	114	89%	.72	22	100%	-	0	-	-	16	94%	.76
Potassium	114	99%	.66	1	100%	-	0	-	-	0	-	-
Blood glucose	114	96%	.68	6	67%	-.20	0	-	-	5	100%	1.0
Blood gases	114	100%	-	0	-	-	0	-	-	0	-	-
Creatinine	114	99%	0.0	0	-	-	0	-	-	0	-	-
ESR	114	100%	1.0	2	100%	1.0	0	-	-	2	100%	-
UA Dipstick	114	92%	.60	8	100%	-	0	-	-	8	88%	.75
ECG	114	90%	.63	12	100%	-	0	-	-	1	100%	-
Joint XRay	114	98%	0.0	0	-	-	0	-	-	0	-	-
CXR	114	97%	.81	10	90%	0.0	0	-	-	4	75%	.5
Mammogram	114	100%	-	0	-	-	0	-	-	0	-	-
Overall	114	75%	.70	86	97%	.38	0	-	-	46	89%	.78

Drug	Agreement on Drug Type			If agree on drug, then agreement on dose and frequency			If agree on drug then agreement on drug-condition contraindications			If agree on drug then agree on drug-drug contraindications		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
ACE Inhibitors	292	97%	.67	10	90%	0.0	10	90%	0.0	10	100%	1.0
Alpha Blockers	292	99%	.50	1	100%	-	1	100%	-	1	100%	-
Beta Blockers	292	96%	.68	13	92%	0.0	13	100%	-	13	100%	1.0
Centrally Acting	292	98%	.59	3	100%	-	3	100%	-	3	100%	-
Calcium Channel Blockers	292	95%	.53	10	90%	0.0	10	100%	1.0	10	100%	-
Direct-Acting Vasodilator	292	99%	.50	1	100%	-	1	100%	-	1	100%	0
Loop Diuretics	292	96%	.71	16	100%	-	16	100%	1.0	16	100%	-
Potassium Sparing Diuretics	292	99%	.50	1	100%	-	1	100%	-	1	100%	-
Thiazide Diuretics	292	99%	.86	10	100%	-	10	100%	1.0	10	100%	1.0
Thiazide + Potassium Sparing	292	98%	.79	10	90%	.74	10	100%	-	10	90%	.74
Nitrate	292	95%	.54	10	100%	1.0	10	100%	-	10	100%	1.0
Insulin	292	99%	.57	2	100%	-	2	100%	-	2	100%	-
Oral Hypoglycemic	292	99%	.74	6	100%	-	6	100%	-	6	100%	1.0
Corticosteroid	292	97%	.59	6	100%	-	6	83%	0.0	6	67%	-.2
NSAID	292	94%	.71	25	96%	.65	24	92%	.80	24	88%	.36
Salicylate	292	92%	.43	11	100%	1.0	11	100%	-	11	91%	.75
Inhaled Beta Adrenergic	292	99%	.66	3	67%	.4	3	67%	0.0	3	100%	-
Cholinergic Inhibitors	292	99%	.50	1	100%	-	1	100%	-	1	100%	-
Xanthine Derivative - Oral	292	100%	1.0	4	100%	-	4	100%	1.0	4	75%	.50
Cephalosporin	292	97%	.75	13	92%	.83	13	92%	.63	13	100%	-

	Agreement on Drug Type			If agree on drug, then agreement on dose and frequency			If agree on drug then agreement on drug-condition contraindications			If agree on drug then agree on drug-drug contraindications		
Erythromycin	292	98%	.68	8	100%	1.0	8	100%	-	8	100%	-
Penicillin	292	99%	.66	2	100%	1.0	2	50%	0.0	2	100%	-
Tetracycline	292	99%	.72	4	75%	.5	4	100%	-	4	100%	-
Aminopenicillin	292	96%	.68	14	100%	1.0	14	100%	-	14	100%	-
Penicillinase-Resistant Penicillin	292	99%	.57	2	100%	-	2	100%	-	2	100%	-
Sulfonamide + Trimethroprim	292	98%	.73	7	71%	-.17	7	100%	1.0	7	86%	.59
Overall	-	-	-	197	95%	.77	196	96%	.77	196	95%	.76

Drugs - Part II - Maryland

Drug	If agree on drug, then agreement on status			If agree on status, then agreement on new drug indications			If agree on status, then agreement on drug monitoring		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Ace Inhibitors	10	90%	.74	2	100%	-	8	88%	.8
Alpha blocker	1	100%	-	1	100%	-	1	100%	-
Beta blockers	13	92%	.63	1	100%	-	12	58%	.33
Centrally Acting	3	100%	1.0	1	100%	-	3	33%	.14
Calcium Channel Blockers	10	100%	1.0	1	100%	-	9	78%	.57
Loop Diuretic	16	88%	.73	9	89%	.61	12	83%	.6
Potassium Sparing Diuretic	1	100%	-	0	-	-	1	0%	-
Thiazide Diuretic	10	100%	1.0	4	100%	-	9	56%	.14
Thiazide + Potassium Sparing Diuretic	10	100%	1.0	2	100%	-	9	89%	.75
Nitrate	10	80%	.6	3	100%	-	3	100%	-
Insulin	2	100%	-	0	-	-	2	50%	.33
Oral Hypoglycemic	6	83%	.67	2	100%	1.0	5	60%	.23
Corticosteroid	6	100%	1.0	4	100%	-	6	83%	.73
NSAID	25	84%	.67	13	92%	0.0	13	85%	.75
Salicylate	11	82%	.63	0	-	-	5	80%	.69
Inhaled Beta Adrenergic	3	100%	1.0	2	100%	1.0	3	100%	-
Cholinergic Inhibitor	1	100%	-	1	100%	-	0	-	-
Xanthine Derivative - Oral	4	100%	1.0	1	100%	-	3	100%	1.0
Cephalosporin	13	92%	.75	10	60%	-.18	0	-	-
Erythromycin	8	100%	-	8	88%	.6	0	-	-

Drug	If agree on drug, then agreement on status			If agree on status, then agreement on new drug indications			If agree on status, then agreement on drug monitoring		
Penicillin	2	100%	-	2	100%	1.0	0	-	-
Tetracycline	4	100%	-	4	100%	1.0	0	-	-
Aminopenicillins	14	93%	0.0	13	85%	.64	0	-	-
Sulfonamide + Trimethroprim	1	86%	0.0	6	83%	0.0	0	-	-
Overall	197	91%	.82	94	88%	.72	107	76%	.62

Prevention - Maryland

	N	% agree	kappa
Blood pressure	106	97%	.39
Flu Shot	106	92%	.82
Mammogram	25	96%	.83
Breast Exam	64	95%	.84

Follow-up - Maryland

	N	% agree	kappa
Consult or Hospitalization	33	97%	.91
Significant Findings	10	100%	1.0

Appendix 2

Diagnoses - Alabama

Diagnosis	Agreement on Diagnosis			If agree on Dx, then agreement on Status			If agree on status then agree diagnosis monitoring			If agree on status then agree on new diagnosis indications			If agree on status then agree on search for treatable causes		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Hypertension	115	98%	.96	42	100%	1.0	40	85%	.74	1	100%	-	1	100%	-
Ischemic Heart Disease	115	100%	1.0	9	100%	-	9	100%		0	-	-	0	-	-
Osteoarthritis	115	97%	.91	29	97%	.87	24	79%	.22	4	100%	-	4	100%	-
COPD	115	97%	.81	7	86%	.72	3	67%	0.0	3	100%	-	3	100%	1.0
Diabetes	115	100%	1.0	15	100%	1.0	13	77%	.67	1	100%	-	1	100%	-
Congestive Heart Failure	115	99%	.94	8	100%	1.0	6	100%		2	100%	-	1	100%	-
Overall	115	96%	.94	110	98%	.91	95	84%	.73	11	100%	1.0	10	80%	.55

Tests - Alabama

Test	Agreement on Test			If agree on test, then agreement on test indications			If agree on test then agreement on test urgent response			If agree on test then agreement on appropriate follow-up		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Blood pressure	86	93%	.79	15	100%	-	0	-	-	5	80%	.55
Hgb/Hct	86	99%	.79	2	50%	0.0	0	-	-	1	100%	-
Cholesterol	86	93%	.74	11	100%	-	0	-	-	11	100%	1.0
Potassium	86	100%	1.0	1	100%	-	0	-	-	1	100%	-
Blood glucose	86	99%	.92	6	100%	-	0	-	-	6	100%	1.0
Blood gases	86	100%		0	-	-	0	-	-	0	-	-
Creatinine	86	95%	.64	4	100%	-	0	-	-	4	100%	1.0
ESR	86	100%	1.0	1	100%	-	0	-	-	0	-	-
UA Dipstick	86	95%	.81	10	90%	0.0	0	-	-	8	88%	.75
ECG	86	93%	.46	3	100%	-	0	-	-	0	-	-
Joint X-Ray	86	97%	.71	4	100%	-	0	-	-	0	-	-
CXR	86	93%	.69	8	100%	-	0	-	-	0	-	-
Mammogram	86	99%	.79	2	100%	-	0	-	-	1	100%	-
Overall	86	78%	.75	67	97%	-.02	0	-	-	37	92%	.83

end

Drug	Agreement on Drug Type			If agree on drug, then agreement on dose and frequency			If agree on drug then agreement on drug-condition contraindications			If agree on drug then agree on drug-drug contraindications		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
ACE Inhibitors	259	100%	1.0	21	95%	.77	6	83%	0.0	21	90%	0.0
Alpha Blockers	259	100%	1.0	4	75%	0.0	1	100%	-	4	100%	-
Beta Blockers	259	99%	.91	5	10%	-	0	-	-	5	100%	1.0
Centrally Acting	259	99%	.91	10	100%	1.0	3	100%	-	10	100%	-
Calcium Channel Blockers	259	99%	.9	15	100%	1.0	5	100%	-	15	100%	-
Loop Diuretics	259	100%	1.0	10	100%	-	4	100%	-	10	100%	-
Potassium Sparing Diuretics	259	100%	1.0	2	100%	-	2	100%	-	2	100%	-
Thiazide Diuretics	259	99%	.87	7	86%	.7	1	100%	-	7	100%	1.0
ACE + Thiazide	259	99%	.80	2	100%	1.0	2	100%	-	2	100%	-
Thiazide + Potassium Sparing	259	99%	.96	13	92%	.63	3	100%	-	13	92%	.81
Nitrate	259	99%	.88	4	100%	-	1	100%	-	4	100%	-
Insulin	259	99%	.93	7	86%	0.0	0	-	-	7	100%	-
Oral Hypoglycemic	259	100%	1.0	8	100%	1.0	1	100%	-	8	75%	.47
Corticosteroids	259	99%	.82	7	100%	1.0	6	100%	1.0	7	100%	1,0
NSAIDs	259	97%	.87	36	100%	1.0	25	76%	.15	36	86%	.46
Salicylate	259	99%	.92	18	100%	1.0	4	100%	-	18	78%	.56
Beta Adrenergic Inhaled	259	100%	1.0	2	100%	1.0	0	-	-	2	100%	-
Cholinergic Inhibitor	259	100%	1.0	2	100%	1.0	0	-	-	2	100%	-
Oral Xanthine Derivative	259	99%	.8	2	100%	-	0	-	-	2	50%	0.0
Beta-Adrenergic Agonist (Oral)	259	100%	1.0	2	100%	-	1	100%	-	2	50%	0.0

	Agreement on Drug Type			If agree on drug, then agreement on dose and frequency			If agree on drug then agreement on drug-condition contraindications			If agree on drug then agree on drug-drug contraindications		
Cephalosporin	259	99%	.97	18	100%	1.0	17	100%	-	18	100%	-
Erythromycin	259	99%	.8	4	100%	1.0	4	100%	-	4	100%	-
Tetracycline	259	100%	1.0	6	83%	.57	5	100%	-	6	100%	-
Aminopenicillins	259	99%	.93	15	100%	1.0	13	100%	-	15	100%	-
Sulfonamide + Trimethroprim	259	99%	.85	9	100%	1.0	7	100%	-	9	89%	0.0
Overall				239	97%	.91	116	94%	.43	239	92%	.62

Drug

Drugs - Part II - Alabama

Drug	If agree on drug, then agreement on status			If agree on status, then agreement on new drug indications			If agree on status, then agreement on drug monitoring		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Ace Inhibitors	21	100%	1.0	6	100%	-	17	82%	.76
Alpha blocker	4	75%	.5	1	100%	-	1	100%	-
Beta blockers	5	100%	-	0	-	-	5	80%	.72
Centrally Acting	10	100%	1.0	3	100%	1.0	9	78%	.68
Calcium Channel Blockers	15	93%	.86	5	100%	-	11	72%	.56
Loop Diuretic	10	80%	.58	3	100%	-	7	86%	.77
Potassium Sparing Diuretic	2	100%	-	2	100%	-	2	50%	0.0
Thiazide Diuretic	7	86%	.59	1	100%	-	2	100%	1.0
ACE + Thiazide Diuretic	2	100%	-	2	100%	-	2	100%	1.0
Thiazide + Potassium Sparing Diuretic	13	92%	.81	3	100%	-	8	75%	.53
Nitrate	4	100%	1.0	1	100%	-	1	100%	-
Insulin	7	100%	-	0	-	-	6	33%	.2
Oral Hypoglycemic	8	88%	.6	1	100%	-	4	25%	-.2
Corticosteroids	7	100%	1.0	6	100%	-	7	86%	.68
NSAIDs	36	97%	.94	24	88%	0.0	12	92%	.87
Salicylates	18	94%	.82	3	67%	0.0	13	69%	.21
Inhaled Beta Adrenergic	2	50%	0.0	0	-	-	1	100%	-
Cholinergic Inhibitor	2	100%	-	0	-	-	0	-	-
Xanthine Derivative - Oral	2	100%	-	0	-	-	2	100%	-
Beta Adrenergic - Oral	2	100%	1.0	0	-	-	2	100%	1.0

Drug	If agree on drug, then agreement on status			If agree on status, then agreement on new drug indications			If agree on status, then agreement on drug monitoring		
Cephalosporin	18	100%	1.0	17	94%	0.0	0	-	-
Erythromycin	4	100%	-	4	75%	0.0	0	-	-
Tetracycline	6	100%	1.0	8	100%	1.0	0	-	-
Aminopenicillins	15	100%	1.0	13	92%	.84	0	-	-
Sulfonamide + Trimethroprim	9	100%	1.0	7	100%	-	0	-	-
Overall	239	96%	.91	110	92%	.82	116	78%	.68

Prevention - Alabama

	N	% agree	kappa
Blood pressure	92	100%	1.0
Flu Shot	90	93%	.85
Mammogram	36	100%	1.0
Breast Exam	59	98%	.96

Follow-up - Alabama

	N	% agree	kappa
Consult or Hospitalization	29	90%	.77
Significant Findings	33	97%	0.0

Appendix 3

Diagnoses - Iowa

Diagnosis	Agreement on Diagnosis			If agree on Dx, then agreement on Status			If agree on status then agree diagnosis monitoring			If agree on status then agree on new diagnosis indications			If agree on status then agree on search for treatable causes		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Hypertension	120	91%	.79	32	84%	.37	25	48	.23	2	100%	1.0	2	50%	0.0
Ischemic Heart Disease	120	94%	.56	5	100%	1.0	4	50%	.2	1	100%		1	100%	
Osteoarthritis	120	88%	.59	27	78%	.50	15	73%	.38	6	83%	.67	5	100%	1.0
COPD	120	90%	.59	11	82%	-.1	9	78%	.5	0			0		
Diabetes	120	94%	.71	10	100%	1.0	9	67%	.54	1	100%		1	100%	
Congestive Heart Failure	120	95%	.71	7	86%	.69	4	100%	1.0	2	100%		2	100%	
Overall	120	77%	.70	92	85%	.54	66	64%	.43	12	75%	.53	11	73%	.42

Tests - Iowa

Test	Agreement on Test			If agree on test, then agreement on test indications			If agree on test then agreement on test urgent response			If agree on test then agreement on appropriate follow-up		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Blood pressure	114	88%	.65	19	89%	-.06	0	-	-	5	100%	1.0
Hgb/Hct	114	97%	0.0	0	-	-	0	-	-	0	-	-
Cholesterol	114	92%	.67	10	70%	-.15	0	-	-	6	67%	-.20
Potassium	114	100%	-	0	-	-	0	-	-	0	-	-
Blood glucose	114	94%	.63	7	100%	1.0	0	-	-	6	100%	1.0
Blood gases	114	99%	.66	1	100%	-	0	-	-	0	-	-
Creatinine	114	94%	.50	4	100%	1.0	0	-	-	3	100%	-
ESR	114	99%	.66	1	100%	-	0	-	-	1	100%	-
UA Dipstick	114	97%	.86	10	90%	.62	0	-	-	9	89%	.78
ECG	114	91%	.53	7	100%	-	1	100%	-	0	-	-
Joint XRay	114	92%	.57	7	71%	-.17	0	-	-	1	100%	-
CXR	114	88%	.30	4	75%	.5	0	-	-	1	100%	-
Mammogram	114	99%	0.0	0	-	-	0	-	-	0	-	-
Overall	104	67%	.62	70	87%	.45	1	100%	-	32	91%	.80

end

Drug	Agreement on Drug Type			If agree on drug, then agreement on dose and frequency			If agree on drug then agreement on drug-condition contraindications			If agree on drug then agree on drug-drug contraindications		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
ACE Inhibitors	297	96%	.54	7	86%	0.0	7	71%	.46	7	100%	-
Alpha Blockers	297	100%	1.0	2	100%	-	2	100%	-	2	100%	-
Beta Blockers	297	97%	.65	8	100%	1.0	8	75%	.5	8	100%	1.0
Centrally Acting	297	98%	.66	5	100%	-	5	100%	-	5	80%	0.0
Calcium Channel Blockers	297	97%	.75	15	93%	0.0	15	93%	0.0	15	100%	-
Loop Diuretics	297	97%	.67	12	100%	-	12	92%	0.0	12	100%	-
Thiazide Diuretics	297	99%	.82	7	100%	1.0	7	86%	0.0	7	100%	1.0
Thiazide + Potassium Sparing	297	99%	.83	10	100%	-	10	70%	-.15	10	100%	1.0
Nitrate	297	96%	.36	4	25%	-.5	4	100%	-	4	100%	-
Insulin	297	99%	.80	4	100%	-	4	100%	-	4	100%	-
Oral Hypoglycemic	297	99%	.85	9	89%	0.0	9	89%	.61	9	78%	.53
Corticosteroids	297	97%	.32	2	100%	-	2	50%	0.0	2	100%	1.0
NSAIDs	297	93%	.64	24	92%	.7	22	86%	.35	22	95%	.78
Salicylates	297	96%	.70	14	93%	.63	13	85%	.41	14	86%	.58
Inhaled Beta Adrenergic	297	99%	.57	2	100%	-	2	100%	-	2	100%	-
Cholinergic Inhibitors	297	99%	.40	1	100%	-	1	100%	-	1	100%	-
Xanthine Derivative - Oral	297	97%	.59	6	100%	-	6	67%	-.2	6	83%	.67

	Agreement on Drug Type			If agree on drug, then agreement on dose and frequency			If agree on drug then agreement on drug-condition contraindications			If agree on drug then agree on drug-drug contraindications		
Oral Beta Adrenergic	297	100%	1.0	2	100%	-	2	100%	-	2	100%	-
Cephalosporin	297	94%	.71	25	96%	.83	25	72%	.19	25	100%	-
Erythromycin	297	98%	.75	8	88%	.71	8	100%	-	8	100%	-
Penicllin	297	99%	.67	1	100%	-	1	0%	-	1	100%	-
Tetracycline	297	97%	.65	8	100%	1.0	8	88%	0.0	8	100%	1.0
Aminopenicillin	297	97%	.73	13	100%	1.0	13	92%	0.0	13	100%	-
Sulfonamide + Trimethroprim	297	97%	.70	10	100%	1.0	10	90%	0.0	10	90%	.74
Overall	297			205	95%	.71	201	85%	.34	203	96%	.83

Drugs - Part II - Iowa

Drug	If agree on drug, then agreement on status			If agree on status, then agreement on new drug indications			If agree on status, then agreement on drug monitoring		
	N	% agree	kappa	N	% agree	kappa	N	% agree	kappa
Ace Inhibitors	7	71%	0.0	0			2	50%	0.0
Alpha blocker	2	50%	0.0	0			1	100%	-
Beta blockers	8	88%	0.0	0			7	71%	0.0
Centrally Acting	5	40%	-.15	1	100%	-	2	0%	0.0
Calcium Channel Blockers	15	80%	.44	2	100%	-	11	73%	.57
Loop Diuretic	12	92%	.83	5	80%	0.0	9	100%	1.0
Thiazide Diuretic	10	100%	-	0	-	-	6	83%	.57
ACE + Thiazide Diuretic	4	100%	1.0	0	-	-	1	100%	-
Thiazide + Potassium Sparing Diuretic	4	75%	0.0	0	-	-	9	67%	.41
Nitrate	9	89%	.73	1	100%	-	1	100%	-
Insulin	2	50%	0.0	0	-	-	2	0%	0.0
Oral Hypoglycemic	9	89%	.73	0	-	-	7	71%	.22
Corticosteroid	2	50%	0.0	0	-	-	1	0%	-
NSAIDs	24	75%	.49	11	45%	0.0	11	36%	.09
Salicylate	14	86%	.71	5	80%	.55	8	50%	.27
Inhaled Beta Adrenergic	2	100%	1.0	1	100%	-	2	100%	-
Cholinergic Inhibitor	1	100%	-	0	-	-	0	-	-
Xanthine Derivative - oral	6	100%	-	0	-	-	5	60%	.17
Xanthine + Beta Adrenergic	2	100%	1.0	1	100%	-	1	0%	-
Cephalosporin	25	92%	.78	18	61%	.09	0	-	-

Drug	If agree on drug, then agreement on status			If agree on status, then agreement on new drug indications			If agree on status, then agreement on drug monitoring		
Erythromycin	8	88%	0.0	7	86%	.59	0	-	-
Penicillin	1	100%	-	1	0%	-	0	-	-
Tetracycline	8	88%	0.0	7	86%	.59	0	-	-
Aminopenicillin	13	69%	.16	8	88%	0.0	0	-	-
Sulfonamide + Trimethroprim	10	80%	0.0	8	88%	.71	1	100%	-
Overall	205	83%	.66	78	72%	.44	88	65%	.48

Prevention - Iowa

	N	% agree	kappa
Blood pressure	115	94%	.33
Flu Shot	111	96%	.90
Mammogram	40	88%	.72
Breast Exam	67	96%	.89

Follow-up - Iowa

	N	% agree	kappa
Consult or Hospitalization	33	88%	.64
Significant Findings	18	67%	-.10

Appendix 4

MEMORANDUM

TO: File
FROM: Ann Lawthers
DATE: November 3, 1993
RE: **Minutes of calls with Alabama, Iowa, and** *Maryland*

GENERAL COMMENTS

Issues with training (length of sessions, time for practice, background and experience of reviewers)	Issues with review implementation (e.g. what workload worked best, fatigue factor, etc.)	Other (e.g. communication with researchers, DFMC)
Future instructions for personnel copying the records should include the importance of capturing the dates of service, visits, calls, tests, etc. *Include vital signs sheets, medication sheets. Ask for initial history and physical, even if outside of time frame, or ask for updates.* *RNs with recent medical/surgical or ambulatory setting experience are best suited for the review. Increased knowledge of current therapies and practices among reviewers could lead to fewer disagreements.* *Manual should be organized by sections flagged, such as RDA*	*High level of concentration rapidly leads to fatigue.* *Weekly short sessions which would highlight unusual events or situations in the ambulatory record would help.* *Do IRR comparisons in a timely fashion so each reviewer will receive feedback that can be applied and increase agreement with other reviewers.*	**Lots of problems with the ID numbers on the sheets. Sometimes had a hard time confirming whether the doctor sent in the correct chart.** **Using HCFA ids and JHU ids was confusing.** **PARADOX - It would be nice to have the capability to flip among screens. in this sense, Paradox was quite unwieldy.** **Flu vaccine and BP checks as the sole reason for a visit do not accurately reflect the care provided and should not be reviewed.**

DIAGNOSES

Diagnosis	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
All Diagnoses	If evidence of treatment or monitoring prior to mention of diagnosis, then it is an old diagnosis. If mention of a diagnosis suddenly appears without symptoms then it is an old diagnosis. Would be helpful to add a help function to system to have lists of synonyms readily available. **Complete hard copies of the screens would have been useful.** **A complete and broad list of synonyms should be included in the manual or on help sheets. The list for the project was quite narrow.** **CAO should by a synonym for COPD, ASHD=IHD, HPT=HTN** **Regular rate is sometimes used as a synonym for pulse. Perhaps a scoring option should include this possibility.**	Always need an N/A. Always need "none of the above" or "Patient not compliant" or "MD ordered and patient didn't follow up." Defaults a problem, may have inadvertently missed something. Can't tell when reviewer missed something, versus a zero score. It would be helpful to look in entire two year record for treatable causes. **Help screens would be very useful.** **Under search for treatable causes, often couldn't distinguish between routine monitoring actions and a workup of a specific diagnosis.**	Often find all except one of the items on the list. Therefore the physician fails. Would be much more helpful if could separate out the various treatable causes and score each separately. The treatable causes screen language is very confusing. Difficult to determine what was OR and what was AND. Much easier to review monitoring actions that were required once than those that were required every six months. For example, if patient has had diabetes for 20 years, then monitoring may be less frequent.	**Severity of illness was an issue for monitoring. Doctors distinguished between the mildly and moderately ill and treated accordingly.**

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Diagnosis	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
HTN	Need to allow inference about diagnosis from meds and lab values, e.g. elevated BP readings and anti-hypertensive meds. These would indicate an "old" diagnosis. Need to establish synonyms for pulse, e.g. regular rate and rhythm or sinus rate. Should allow synonyms of Hptn or Hpn.		MONITORING: The diet criteria may not apply. If HTN is mild and diet well controlled than physician may not mention. Also diet might work better if reviewed once rather than 6 monthly. Blood pressure and pulse should be monitored separately.	Some physicians put vital signs on a separate sheet, maybe we didn't get the sheet when we requested the records. Need to make sure it is requested. Need to request any hospital admission H & P. This is where we might find the breast exam and the fundoscopy.

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Diagnosis	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
DM	Need to allow inference about diagnosis from meds and lab values, e.g. elevated glucose levels and anti-diabetic meds. These would indicate an "old" diagnosis. **Need to able to review if patient on oral hypoglycemics, even if diabetes is never mentioned.**		NEW DIAGNOSIS: Too narrow to be used in 2 year review period. Criteria also difficult to meet if diagnosis was made during a hospitalization. TREATABLE CAUSES: If on steroids, this criteria really needs a response #9, not applicable.. MONITORING: HbgA1C rarely done every 6 months. Comment on blood glucose control and noting the blood glucose level in MR seemed redundant.	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Diagnosis	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
COPD	Additional synonyms are required, e.g. chronic lung disease, obstructive lung disease, silicosis.		NEW DIAGNOSIS: Should allow XRay findings of COPD to be allowed for a new diagnosis. **Productive cough lasting 3 months for 2 consecutive years is impossible to evaluate given the 2 year review window.** Evidence of PFT difficult to interpret following the wording of the criteria. TREATABLE CAUSES: H/O tobacco use or environmental precipitant may be difficult criteria to meet because of limited time frame. Also can't tell who the non-smokers are because question about smoking may be H&P done years ago. MONITORING: Same problem with smoking as noted above. Separate Pulse and respiratory rate.	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Diagnosis	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
OA	Degenerative arthritis should be allowed as a synonym of OA. Some records contained evidence of both rheumatoid and osteoarthritis. It became difficult to tell which form of arthritis was being discussed yet just the word arthritis got one into the system. **This review was confusing. Many times OA was an incidental finding. The monitoring need not apply to mild disease.**		TREATABLE CAUSES: Obesity is not necessarily a treatable cause. Height measurements were often absent. Many doctors do not repeat height on adult patients after initial visits. As review progressed, interpretation of criteria changed. If it was very clear because of the patient's weight alone, that the patient WAS or WAS NOT obese, credit was given for search for treatable causes. Should be clarified. MONITORING: Side effects of medication needs to be clarified. Any medications, including those which we do not review (e.g. Gold, narcotics, muscle relaxants, forms of acetaminophen)? Any specific side effects (e.g. GI discomfort, bleeding, renal, or liver)? Any specific monitoring activities (e.g. lab, hemoccult, urinalysis, GI check)?	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Diagnosis	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
IHD	If evidence of "ischemia" and or S/P CABG, this should be enough evidence to enter IHD as a diagnosis.		MONITORING: Very reviewer friendly. Problem with smoking criteria, can't tell who is a non-smoker. NEW DIAGNOSIS: Criteria were very narrow. Often there was evidence of "ischemic changes" on an EKG but this could not be worked into existing criteria.	
CHF	Manual and criteria shown on screen are inconsistent regarding thyroid disease.		NEW DIAGNOSIS: Took a really through chart search to meet criteria for a new diagnosis. TREATABLE CAUSES: All of listed criteria were extremely difficult to meet in the reviewed medical records. Usually absent item was "statement regarding presence/absence of thyroid disease."	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

TESTS

Test	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
All Tests	Need a special section in the manual which covers the values which define renal disease, chronic renal insufficiency.	Always need an N/A. Always need "none of the above" or "Patient not compliant" or "MD ordered and patient didn't follow up." Defaults a problem, may have inadvertently missed something. Can't tell when reviewer missed something, versus a zero score. Would have worked better to indicate any abnormal result the first time it is encountered and then say that treatment is underway. The time frames before the test was reviewed didn't work very well, particularly for cholesterol, creatinine, and ESR. Time frames particularly troublesome when pt had frequent hospitalizations. **Recording the test value would have been useful.** **The way the screens were set up, the doctor was flagged if he/she missed just one in a long complex sequence of actions.**	If physician had a plan, even if it wasn't what was on the screen, it was scored "7", to indicate something was being done.	Often found incidental or "red herring" abnormal levels in panel tests. What to do with them?

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Test	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
BP		Big problem with "new finding", should only be reviewed when find two abnormal yet the screens get you into the system with a single abnormal. Physicians flagged because of lack of scoring options. Couldn't score N/A. Screens inconsistent. Is it > 90 of >= 90? Should it be 91 - 104?	**Isolated systolic hypertension needs to be addressed with separate criteria.**	
Cholesterol		**The screen for cholesterol, in particular, didn't work well. The screen instructs the reviewer to look ahead, yet if the first noted abnormality is in the middle of a stream of abnormals, the appropriate work-up may be in the past.** **Test indications should include personal as well as familial history.**	**Need a separate set of criteria for "old" cholesterol abnormalities. Even with the time frame restriction, still ended up reviewing cholesterol values that were in the middle of an abnormal stream, where the initial corrective actions were taken before the review period.**	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Test	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
Potassium			Test result of greater than 6 poses a problem unless it meets the criteria for urgent response. There is no way to review such a result (and there should be).	
Creatinine	Definition needs to be in manual in a special section.	It would be very helpful to have an alt screen to define renal disease.		
ESR		Screen really needs a NA option	Test response does not have a choice for "ESR of 30 to 80" or greater than 80 with cause known or symptoms of inflammatory arthritis.	
ECG			Routine screening should be a justification for ECG.	
Joint XRay	**Wording on what is abnormal needs to be more definitive.**	**There was no way to take action other than hospitalization for the urgent response screen.** **The 2nd test response range, "normal x-ray after trauma" can never be selected. DEMPAQ does not review normal findings.**		

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Test	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
Mammogram			Options for test response should include "suspicion" of malignancy.	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

DRUGS

Drug	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
All Drugs	Propose that only drugs in the first visit be considered old drugs. All the rest would be new. A note indicated that the drug was being d/c'd indicates it is an old drug. In some instances, there is no clue as to whether the patient remained on the drug. If don't see d/c'd then assume the drug was continued. If another medication of same type was substituted then assume the first drug was d/c'd [Note from Ann, should still continue to monitor though.] **Old and new drugs a problem. Sometimes could tell form telephone contacts whether the drug was new or old, but telephone contacts weren't reviewed. Suggest, for medications only, reviewing drugs which are initially prescribed without a visit.**	Always need an N/A. Always need "none of the above" or "Patient not compliant" or "MD ordered and patient didn't follow up." Need option, "Monitoring done but not within recommended time period." There is often a problem with office scheduling. Can't get appointment and therefore the physician is in non-compliance. Also suggest changing to allow credit for test being done sometime "within" the recommended time period, not necessarily at the 3, 6 month mark. Defaults a problem, may have inadvertently missed something. Can't tell when reviewer missed something, versus a zero score. Liver impairment as a contraindication was problematic because did not have criteria to follow.	Need to allow for other indications found in the PDR. Physicians get flagged when condition not on approved list, yet the condition is in the PDR. Need to allow inference regarding diagnosis indications for a drug. E.G. patient may have BP readings indicative of HTN but physician never says HTN, yet prescribes an HTN drug. Absolute contraindications were difficult to review. Should be allowed to look for contraindications after the initial prescription. [Ann note, I thought we did]	Big problem with drugs. Nurse in the office needs to list the medications the patient is on. Knowing what pt is taking in total is problem that needs to be addressed by the physician community. All meds need to be listed, not just those prescribed by this doctor.

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Drug	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
	Nursing home visits skew old and new drugs. For example, the patient sees the doctor at the nursing home, prescribes a drug. Several months later the patient is seen in the office and it is the first mention of the drug in the record, yet it is not a new drug. **The manual should list all the drugs, generic and trade, and then page number reference.** **Stepped therapy is confusing. The patient may have completed step 1 but can't really tell.**	**Need to have an option that monitoring is responsibility of clinic or consultant.** **The contraindications should be separated out by item.** **It is not clear for the risk-benefit parameter, whether the alt screens apply to the general category or the specific drugs (e.g. reserpine or epinephrine).** **Illegible should be added to the monitoring screen scoring options.**	**Consider adding Keftabs and Cipro.** **Also the following medications were identified during the review but were not on the F1 listing (and probably should be):** **Vanceril Cardura** **Beclovine Septra DS** **Anacin EES** **Nitropast Keftab** **Triamterene with HCTZ** **Cipro Robitet** **Azamcort Inhaler** **PCE Cleocin** **Floxin**	
Loop Diuretic	If diuretic drug started on a course and then changed to PRN, should not subject to monitoring. Need a manual and criteria change.		Lasix issue, CHF not listed. Physician may not list edema as problem, yet if pt. has CHF, this should be an allowed indication.	
Potassium Sparing Diuretic	If diuretic drug started on a course and then changed to PRN, should not subject to monitoring. Need a manual and criteria change.		**The criteria should allow for the situation where a K+ supplement is added to a K+ sparing diuretic when the K+ is low**	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Drug	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
Thiazide Diuretic	If diuretic drug started on a course and then changed to PRN, should not subject to monitoring. Need a manual and criteria change.			
Thiazide + K+ sparing Diuretic	If diuretic drug started on a course and then changed to PRN, should not subject to monitoring. Need a manual and criteria change.			
Nitrate			**It is not recorded whether it is used intermittently or continuously.**	
Insulin	**When both humilin and NPH insulin are prescribed, which should be reviewed.**		**In Iowa have many situations where both oral hypoglycemics and insulin are given. The criteria do not address this situation.**	
NSAID			**Prophylactic aspirin should not be contraindicated with an NSAID. Need to discuss the issue of dosage when determining contraindications.**	
Salicylate	Should review prophylactic aspirin. Can still lead to problems that should be monitored.		**Aspirin prophylaxis complicates matters.**	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Drug	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
Xanthine-Derivative			If patient does not have weight change, nor symptoms then probably wouldn't monitor theophylline levels. **The time frame for theophylline levels was much too rigid.**	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Drug	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
Cephalosporin		**When the reviewer pulls down the alt screen for indications, the text reads "More indications," however, there are no prior indications.**	Criteria need to be re-examined. Does not include enough indications. The instruction look in entire MR for indications is inappropriate. For example, the doctor might prescribe Bactrim for symptoms of an unspecified type of upper respiratory infection (e.g. cough, sore throat) which is really not enough of an indication for this medication. Yet, looking in the entire MR, the patients has a history of UTI's. Shouldn't the reviewer just consider the situation for which the antibiotic is now prescribed? **Prescribing a cephalosporin in the presence of a penicillin allergy is not necessarily a contraindication if the original reaction was mild.** **Also, requiring cephalosporin to be a drug of second choice (after penicillin) can waste a lot of the patient's time.**	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Drug	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
Erythromycin			Criteria need to be re-examined. Does not include enough indications. See note above about looking in entire MR. ? indicated for mycoplasma without allergy to penicillin **Pertussis as an indication for the elderly doesn't make sense. The incidence of pertussis in the elderly is negligible.**	
Tetracycline			Criteria need to be re-examined. Does not include enough indications. See note above about looking in entire MR. Also ? indicated for diverticulitis.	
Aminopenicillin			Criteria need to be re-examined. Does not include enough indications. See note above about looking in entire MR.	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Drug	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
Penicillin and others			Criteria need to be re-examined. Does not include enough indications. See note above about looking in entire MR.	
Sulfonamide			See note above about looking in entire MR.	
Bactrim			See note above about looking in entire MR.	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

PROCEDURES

Procedures	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
All Procedures		Always need an N/A. Always need "none of the above" or "Patient not compliant" or "MD ordered and patient didn't follow up." Need "abnormality not listed" **Screens were very difficult to work through. Need to be revised.** **It would be clearer if cecum, right, transverse and descending colons each had a response number.**		

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

FOLLOW-UP

Procedures	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
All Follow-up		Always need an N/A. Always need "none of the above" or "Patient not compliant" or "MD ordered and patient didn't follow up." **Perhaps "illegible" is an option for scoring.**		
Mass, Lump, Bleed, Weight Loss	Need to be clear in manual that should not include patients on a weight loss program. The criteria for including in the review should be "significant for the patient." Need to be more clear about when to review sign. Should they be reviewed multiple times if they appear in the record multiple times? Another problem with lesions is what qualifies as a lesion. Is this an ulceration (diabetic) or burn or traumatic injury or only a potential malignancy. What types of bleeding should be reviewed? Frank bleeding? Microhematuria? RBC 2-3, RBC 8-10 or TNTC in urinalysis? **Confusing what kind of bleed to review.** **Lesion can refer to so many things. Skin lesions, are they lumps? Is the review most interested in potential malignancy? The manual should be crystal clear about what the review is really looking for.**			

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Procedures	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
Hospitalization/ Consultation	**What does "Patient *is* or will be admitted mean? The review does not cover hospitalizations. Perhaps the manual needs definitions.**		Need to include option of patient refusal. Also, doctor ordered.	

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

PREVENTION

Prevention	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
All Prevention		Always need an N/A. Always need "none of the above" or "Patient not compliant" or "MD ordered and patient didn't follow up." In many instances, it is clear from the record that the physician ordered the preventive care yet no evidence that the patient received it. **This review should come at the end of the review rather than the beginning. After moving back and forth through the chart in the course of the review, have a better handle on content of chart.**		
Mammogram	Need to make clear in the manual that actual documentation of care being given is required for mammogram.			Flags being generated for men. Issue with scoring program.
Breast Exam	Not clear how to review breast exam.			Flags being generated for men. Issue with scoring program.
Flu Shot	Need to make clear in the manual that actual documentation of care being given is required for flu shot.			

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

RDA

RDA	Problems with manual, e.g. instructions regarding what to review, definitions (e.g. abnormal LFTs, renal disease, etc.), old versus new diagnosis	Problems with Screens, e.g. wording, scoring options	Problems with review criteria	Other
All RDA	Need to include explicit decision rules in special tabbed section.	Problem is difference between 1 and 2. Would suggest expanding the scoring to 1-3 with 4 as illegible. The 1 - 2 scale is not sensitive enough to capture differences in documentation.		

Data are for Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Chapter 14
Evaluating the DRRS: Assessment by Physicians

Abstract
The DEMPAQ project used a structured peer review process to test the validity of performance measures designed to assess the quality of ambulatory care delivered to Medicare patients in three states. The peer review procedure described here does not confirm that a quality problem exists, rather it aims to validate the performance measures (including the criteria, the data sample, the data abstraction and decision rules used for the medical record review).

Methods
Medical records data were scored using a SAS program which calculated rates of acceptable performance on each of 22 indicators. Cases which apparently did not meet the review criteria were identified as "flags." We randomly sampled six flags per indicator, per item, per state, for the peer review and reviewed approximately 263 different criteria in each of three states.

The peer review instrument consisted of a two-page form and was administered in a nurse-facilitated review setting. The first page of the review form described the flag and presented the questions to be addressed by the physician reviewer. The second page of the form displayed the computer screen and scoring options, as seen by the nurse reviewer, and the specific criteria that were used to evaluate the item. Prior to meeting with the physician peer reviewers, the nurse reviewer examined the chart to become familiar with its contents. During the review, the nurse reviewer's role was to answer questions, direct the physician reviewer to pertinent sections of the chart and to make sure that the review form was properly completed.

Results
As an illustration, we present the results for a single clinical item, hypertension (complete data may be found in Appendix 2). The original review showed a flag rate of 48% for the monitoring indicator. In the validity review, the flag was upheld 12 out of 19 cases (63%) suggesting that only these cases were truly quality problems. Revisions to the criteria were recommended in four cases and patient factors explained two of the flags. Overall, the data suggest that the true flag rate for old diagnosis monitoring is 32% rather than the original score of 48%.

Conclusion
We conclude that a structured peer review procedure is a useful way to measure validity and produces results that lead to improvements in the performance measures and also aid in interpreting the findings of the review.

This chapter describes a structured peer review process which tests the validity of the performance measures designed for the DEMPAQ Record Review system (DRRS). Earlier, Chapter 10 presented the results of the validity studies for the claims profiling. Some of the background discussion about validity is repeated here, under the assumption that some readers will be interested in the record review only.

We use the peer review process to validate both the review criteria and the abstracting rules for the specially designed computerized review system. We also use the peer review process to set rough confidence limits around the results. This use of peer review represents a departure from the traditional purpose of peer review, which has been to confirm quality problems.

In this chapter, we first describe the questions to be evaluated by a validity study of the record review criteria. Next, we outline the framework and steps for the peer review process, followed by a discussion of the study's results. We conclude with a discussion of the implications of the results for performance measure development.

Research Questions

The notion of validity encompasses a variety of concepts. Face validity is whether the measures appear to make sense (Abramson, 1990), i.e. how well does a performance measure appear to measure the quality of care? Criterion validity is how well the results of the measurement compare to a gold standard (Abramson, 1990). While a gold standard for quality does not exist, the concept of comparing the interpretation of the profiling results against an interpretation derived from another data source proves useful. This process of comparison allows us to identify various factors which affect the interpretation and validity of the record review criteria.

We wished to answer two questions with the DRRS validity study:

- How valid are the criteria used to review the office record?
- How valid are inferences about the quality of medical care received by a group of patients, based on performance rates derived from individual clinical items? How often do physician reviewers agree with a flag? What patient, practice or other factors that may be discovered in the office record explain why a particular clinical item is flagged?

Our peer review is different from traditional peer review. Instead of confirming that a quality problem exists, the DRRS peer review process is designed to improve the way in which quality of care reviews are performed. We are primarily interested in evaluating and validating the criteria, not the doctor being reviewed. A second goal of the peer review process is to quantify our confidence in the results of the record review.

Methods

As described in Chapter 11, the record review was carried out using a computerized data abstraction instrument which permitted nurse reviewers to enter information directly into computer databases. At the completion of the review period, the data were downloaded from Paradox to text files which could then be read by a scoring program written in SAS. The scoring program created a score for each indicator, by clinical item. Scoring permitted us to categorize the results of the review. Possible scores were "met," "flag," "acceptable alternative," "not applicable" and "not reviewable." Indicators considered not applicable and not reviewable were not scored. A score of "met" means that the abstractor found evidence of compliance with the review criteria in the medical record. Conversely, a "flag" score indicates that the review criteria appear not to be met. An "acceptable alternative" is a sound reason why a recommended action could not or should not be done. "Not applicable" means the criteria do not apply, and "not reviewable" indicates the time frame for records data was insufficient to review the case.

Peer Review Sample

Peer review was conducted on a sample of cases where an episode of care had been flagged for an apparent failure to meet clinical criteria. The sampling was conducted in two phases: the first when approximately one-third of the review had been completed and the second when the review was nearly complete. Thus the cases for peer review came from the first third of reviewed cases and the last two-thirds.

The sampling frame consisted of all flagged cases. The flags were stratified by indicator and clinical item. Within each indicator/clinical item cell, a random sample of three flags was taken. Since the sampling took place in two phases, the total number of flags reviewed per indicator/clinical item was six. Each indicator/clinical item cell corresponds to a single criterion, with the exception of the drug and diagnosis monitoring indicators. These indicators were calculated from a maximum of eight individual criteria.

Peer Review Form

The peer review form consists of two pages: the first shows the review questions and a self-adhesive, computer-generated label identifying the case and flag to be reviewed; the second displays the criteria specific to the flag listed on the label. The entire peer review form is shown in Appendix 1.

The label at the top of the first page is generated by computer with all the relevant information necessary to identify the case and flag to be reviewed. Below the label is a list of possible evaluations of the flag. After reviewing the patient record and applying criteria from page two, the physician reviewer is asked to select from among seven possible responses the answer which most closely fit his/her assessment of the flag. Additional space is included at the bottom of the page for explaining the response.

The physician reviewers were asked to consider the following issues in reviewing the patient record:

- Are the review criteria appropriate for the indicator?
- Are the review criteria applicable to the case at hand?
- Was the flag justified in this case?

The second page of the peer review form was specific to the function/item/flag combination being reviewed. The format of the DRS question was reproduced at the top of the second page as it appeared on the computer screen used by the initial nurse reviewer to score the indicator with the text of the criteria used for review of that item reproduced below. By seeing the actual review question which appeared on the data entry screen, and the list of possible responses, along with the clinical criteria unique to that item, the physician reviewer reconstructs the rationale for the original review which resulted in a flag. This assisted the physician reviewer in assessing the validity of the flag. If the physician reviewer determined that revisions to the criteria were appropriate, he/she was asked to write his/her proposed revisions on the first page of the form, along with their rationale for the change.

Review Procedure

As a first step, the nurse reviewer placed the self-adhesive label on the review form and selected the appropriate second page based on the flag and clinical item shown on the label. Next the nurse reviewer examined the chart to become familiar with its contents and to try to reconstruct how the flag occurred. During the physician review sessions which followed, the nurse reviewer was available to answer questions about how the review was conducted, to assist the physician reviewer in finding data in the chart and to ensure that the review form was correctly filled out.

Results

The complete results for all clinical items are shown in Appendix 2. The results are grouped by function. Below, the Maryland results for a single clinical item, hypertension, illustrate how the data from this process can be used to modify the interpretation of the record findings and to undertake ongoing revision of the criteria.

Hypertension Monitoring

The indicator, appropriate monitoring of a preexisting diagnosis, is composed of eight separate criteria. If 60% or more of the criteria are met then the indicator receives a met score; if less than 60%, the indicator receives a flag. For hypertension, 48% of 271 cases were scored as having flags.

A total of 19 cases were reviewed for flags on hypertension monitoring from the three study states. Of the 19 cases, the flag was upheld in 12 instances. In one case the nurse reviewer had made an error in abstracting the material. For two cases, patient factors meant that the criteria could not be met. In the remaining four cases the physician reviewers suggested revising the criteria. Three of the physician reviewers suggested that time frames for monitoring were too rigid. The criteria called for monitoring diet, drug compliance, pulse and blood pressure every six months. By definition, the reviewer's manual required that this time frame be applied as every six months with a leeway plus or minus two weeks. The proposed revision would permit nurse reviewers to accept monitoring done within a six-month period rather than looking for evidence of monitoring at six-month intervals. The fourth physician reviewer questioned the need for a fundoscopy exam by the primary care physician. He suggested that primary care physicians are not particularly adept at fundoscopic exams. Furthermore, he questioned whether the treatment given by a primary care physician would be changed as a result of the fundus exam done in the office.

The above results for monitoring have the following uses. First, 63% of the flags were upheld (12 out of 19). This implies that 63% of the cases scored as flags are probably valid or true quality problems. Thus the true flag rate for old diagnosis monitoring is 30% of all cases (.63 x the original rate of 48). Second, the criteria may be revised based on the suggestions of the physician reviewers. For example, the issue of time frames appeared repeatedly throughout the peer review process. One of three steps may be taken to correct the problem. First, the abstracting rule may be changed to permit actions found within the time period being scored as acceptable performance. Second, the plus or minus interval around the time frame may be widened, e.g., from 6 months + 2 weeks to 6 months + 2 months. Third, a scoring option may be added to the screens "Monitoring performed, but not within specified time frame."

Hypertension New Diagnosis

Hypertension as a new diagnosis has two indicators: new diagnosis based on appropriate evidence and search made for treatable causes/comorbidities of new diagnosis.

Fourteen cases were peer reviewed for the evidence indicator. Of these, only six flags were upheld as valid. This reduces the percentage of cases flagged for lack of evidence from 27% to 12%. In one case, the abstracting instructions were imprecise: in one case extenuating circumstances prevented the criteria from being met; in four cases the nurse reviewer had incorrectly extracted information; and in two instances, the physician reviewers suggested revising the criteria.

Nineteen cases were reviewed on the treatable causes indicator. Of these, seven of the flags (37%) were upheld. Abstracting instructions needed revision in five of the cases and criteria required revision in six cases. One case did not meet the criteria due to extenuating circumstances.

Discussion

The practice of medicine is continually changing. As new technologies and treatments revolutionize medicine, providers must learn when and how to use them appropriately. New technologies are often expensive and potentially dangerous; using them appropriately and implementing them safely is an issue of quality. Clinging to outdated, ineffective and more dangerous treatments in the face of newer, better alternatives is also an issue of quality.

To keep up with the changes in medicine, methods to monitor and improve quality must also change. As new methods and tools appear, they must be validated in order to be useful. The approach to validating performance measures presented here has several important advantages. First, because the process is structured, there is consistency from review to review. Physician reviewers use the same structure for classifying problems with the criteria, regardless of the time of the review, the state in which the review is being conducted, or the medical background of the physician reviewer. This permits us to draw reasonable comparisons and conclusions from the data.

Second, the data are easy to analyze because the categories of responses are predetermined. We can identify precisely how many cases had patient or practice factors mitigating the flag. Third, we capture major categories of acceptable alternatives: patient; practice; and extenuating circumstances. This helps in both fine tuning the criteria and in "adjusting" for factors known to affect the results. Fourth, the process generates explicit changes to criteria rather than simply noting "this needs to be changed." Because the suggested changes are based on actual patient situations, the changes are clinically meaningful. Finally, physicians report that they "learned something" from doing the review. This led us to design a CME workshop around the structured case review process used during the project's peer review phase.

There are, of course, limitations to our structured approach to peer review. First, it is likely to be more expensive than traditional peer review because the process involves a nurse reviewer facilitating the physician reviewer process. Second, it would be prohibitively expensive to review all the flags for validity; therefore only a sample is scrutinized by physician reviewers. Over time, the strategy for conducting the structured peer review would be to periodically review small, focused samples of cases to calibrate the explicit review criteria. This strategy minimizes the cost of the peer review while achieving the goal of constantly updating the system.

Our vision for this type of process involves multiple iterations of review and feedback. As flags are reviewed and the criteria revised, the validity of the review criteria improves. Over time, the changes required to calibrate the criteria become smaller and smaller until a steady state is reached. At this point, only a change in what is considered appropriate medical practice will alter the criteria. When such a change occurs and the criteria are revised, then the whole process of calibration begins anew.

Summary and Recommendations

We have described a structured approach to peer review that has two goals: to validate review criteria and provide rough confidence limits on the interpretation of results. We believe this approach has significant advantages over traditional forms of peer review and can be used as a model for other validation efforts.

Should the DEMPAQ record review tool be used for further pilot testing or perhaps national implementation, the following steps must be taken:

- Review physician recommendations for criteria changes (see Appendix 2) and determine which changes to implement; and
- Revise criteria.

Appendix 1

DEMPAQ PEER REVIEW WORKSHEET

Peer reviewer: ________________
Date: _____
PRO: _____

PATIENT ID: 0014422
FLAG: D2
CODE: 2

FLAG SCORED FOR:
DIAGNOSIS: Ischemic Heart Dis on 11/11/1988

REASON:
New diagnosis not justified by evidence in medical record.

The patient care episode shown on the label above was flagged for physician review because it failed to meet the DRS review criteria. On the attached page you will find the review format and the criteria used by the initial PRO reviewer. Please review the medical record for the item of care shown above, and indicate below your own judgement regarding the item in this case.

CONSIDER: **Are the review criteria appropriate for this indicator?**
Are the review criteria applicable to the case at hand?
Was the flag justified in this case?

Choose only the one response which most closely fits your opinion. Use the spaces below and the back of this page to give your comments.

__ 1. The criterion flag was justified given the data in the medical record.

__ 2. The criterion could not be met in this circumstance because of patient factors. Describe the patient factors below under comments.

__ 3. The criterion could not be met in this circumstance because of practice factors. Describe the practice factors below under comments.

__ 4. The criterion should not be met in this case due to extenuating circumstances. Describe the extenuating circumstance below under comments.

__ 5. The flag is not justified because of abstractor error. Describe the error under comments.

__ 6. The flag is not justified because the abstracting decision rule is incorrect or incomplete. Suggest a revised decision rule below.

__ 7. The flag is not justified because the criterion needs to be revised. Propose a revised criterion for this item below.

CRIT/CODING RULE REVISIONS:____

COMMENTS:____

CONTINUE COMMENTS ON THE BACK OF THIS SHEET

Is the documentation in this record adequate to conduct a valid quality of care review? ____

Ischemic heart disease

FLAG:D2
CODE: 02

Review question displayed on DRS screen

Evidence for a New Diagnosis

Patient Id: Diag.: Visit date:

Indicate if any of the following evidence for a new diagnosis is in any note in the medical record: 1=yes 0=no

Look in entire MR.

0◂

Clinical criteria displayed in window on DRS screen:

EVIDENCE ESTABLISHING A NEW DIAGNOSIS:

1. Cardiac chest pain (worse with exercise, relieved by rest) with ST segment changes on ECG
   OR
2. Positive exercise tolerance test
   OR
3. Positive cardiac catheterization

Appendix 2

DIAGNOSIS ANALYSIS
Peer Reviewer Comments, Three State Combined Data
November 3, 1993

HYPERTENSION **CODE: 01**

Indicator	Criterion	Peer Review
Old Diagnosis Monitoring	Overall	Flag justified = 12 Manual revision = 1 Patient factors = 2 Criteria revision = 4 Total flags reviewed = 19
	Monitor diet - q 6 mos	1-3. *It should be acceptable to do monitoring within a 6-mo period, not q 6 mos.*
	Fundoscopy - once	**4. Do not think it is necessary to check the fundus to assess BP control. It is not a technique that many are very good at. Treatment would not generally be changed as result of fundus exam.**
Evidence for a new diagnosis	DBP ≥90 on two occasions OR SBP ≥140 on two occasions OR any DBP ≥130	Flag justified = 6 Manual revision = 4 Extenuating circumstances = 1 Decision rule = 1 Criteria revision = 2 Total flags reviewed = 14 1. MD is indicating a tentative diagnosis; notes that he will repeat BP and does so. 2. Borderline HTN; does not fit criteria.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Search for treatable Causes	If DBP >105 on two occasions, look for renal dis. AND hyperaldosteronism AND obesity AND inquire re alcohol intake AND if on hypertensive med., possible causative role.	Flag justified = 7 Manual revision = 5 Extenuating circumstances = 1 Criteria revision = 6 Total flags reviewed = 19 1. Criteria does not include elevated diastolic BP <105. 2. Criteria does not allow for diagnosis to be made based on evidence of treatment. 3. Criteria for evaluation of DBP >105 not applicable to a patient with DBP 80-98. Need to change scoring. 4. Diastolic never ≥105. Need to revise criteria. **5. It should not be necessary to document all negative criteria.** **6. Sometimes need to exercise caution in documenting social problems in chart, such as drug/alcohol use; therefore, eliminate alcohol use. Should add a criterion to look for end-stage damage of HTN.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DIAGNOSIS ANALYSIS
Peer Reviewer Comments, Three State Combined Data
November 3, 1993

ISCHEMIC HEART DISEASE **CODE: 02**

Indicator	Criterion	Peer Review
Old Diagnosis Monitoring		Flag justified = 5 Manual revision = 2 Extenuating circumstances = 1 Total flags reviewed = 8
Evidence for new diagnosis	Cardiac chest pain OR exercise tolerance test OR positive cardiac catheterization	Flag justified = 10 Manual revision = 3 Decision rule = 1 Criteria revision = 3 Total flags reviewed = 17 1. Criteria needs to be expanded to allow therapy based on reasonable suspicion. 2. Criteria should be expanded to include evidence of silent ischemia in diabetics. **3. Exercise-induced cardiac pain needs to be added.**
Search for Treatable Causes	HTN AND valvular heart disease AND anemia AND inquire re smoking habits AND obesity AND serum cholesterol AND blood glucose	Flag justified = 5 Manual revision = 5 Criteria revision = 2 Total flags reviewed = 12 1. Criteria should be expanded to allow history of MI as ischemic heart disease. 2. *Unreasonable criteria - any one of co-morbidities would be adequate.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DIAGNOSIS ANALYSIS
Peer Reviewer Comments, Three State Combined Data
November 3, 1993

OSTEOARTHRITIS **CODE: 03**

Indicator	Criterion	Peer Review
Old Diagnosis Monitoring		Flag justified =10 Manual revision = 4 Extenuating circumstances = 2 Decision rule = 2 Total flags reviewed = 18
Evidence for new diagnosis	Symptoms lasting longer than 1 mo that are suggestive of osteoarthritis OR positive x-ray findings of osteoarthritis	Flag justified = 7 Manual revision = 4 Extenuating circumstances = 1 Decision rule = 3 Criteria revision = 4 Total flags reviewed = 19 Decision rules 1. Criteria does not give strict definition to deal with probable diagnosis. **2. Seems to be an old diagnosis, even though this was first mention of symptoms.** **3. Physician states "she has some DJD." This indicates a history of OA and, therefore, is an old diagnosis.** Criteria revisions 1. Two-yr time frame is too limited for diagnosis of osteoarthritis. *2. Incidental findings without symptoms need not be addressed.* **3. We cannot expect system to justify every comment made in record. In this case, a complaint of arthritic pain does not equal a diagnosis of arthritis.** **4. Add "history of OA" to criteria.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DIAGNOSIS ANALYSIS
Peer Reviewer Comments
November 3, 1993

CHRONIC OBSTRUCTIVE PULMONARY DISEASE

CODE: 04

Indicator	Criterion	Peer Review
Old Diagnosis Monitoring	Pulse, respiratory rate - once	Flag justified = 6 Manual revision = 7 Patient factors = 1 Criteria revision = 4 Total flags reviewed = 18 **1. Would not include RR unless symptoms are present.** **2. Would separate pulse and RR into separate criteria.**
	Chest x-ray - once	*3. CXR not indicated unless symptomatic; should consider PFTs.*
	Tolerance of exertion - once	*4. Comments on shortness of breath may supersede tolerance of exertion.*
Evidence for New Diagnosis	Productive cough lasting 3 mos for 2 consecutive yrs AND dyspnea on exertion with or without wheezing OR decreased forced expiratory volume (1 second)	Flag justified = 12 Manual revision = 1 Decision rule = 2 Total flags reviewed = 19 **1. Should allow COPD dx given DOE with increased expiratory phase by lung exam. Patient refused PFTs, ABGs.** **2. Need to allow more synonyms for COPD.** 1. Criteria needs to include chest x-ray findings of COPD as evidence supporting diagnosis.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DIAGNOSIS ANALYSIS
Peer Reviewer Comments, Three State Combined Data
November 3, 1993

DIABETES **CODE: 05**

Indicator	Criterion	Peer Review
Old Diagnosis Monitoring	Overall	Flag justified = 10 Manual revision = 1 Extenuating circumstances = 1 Practice factors = 1 Patient factors = 2 Criteria revision = 3 Total flags reviewed = 18
	Weight, pulse, BP - q 6 mos	**1a-2a-3a. Pulse should be deleted.**
	HgbA1C or fructosamine - q 6 mos	**2b. Only when symptoms show patient is less controlled should HgbA1C be done q 6 mos.**
	Urinalysis dipstick for protein and microscopic exam of sediment - q 6 mos	*1. Should be acceptable to do within 6 mos.* **1b. U/A should be done once/yr.** **2c-3b. Only when symptoms show patient is less controlled should U/A be done q 6 mos.**
	Funduscopy - once	**2d. Not useful in primary care because cannot see enough to say if problem.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Indicator	Criterion	Peer Review
Evidence for a New Diagnosis	Blood glucose >250 with symptoms of DM OR two FBS >140 OR blood glucose levels during GTT >250 at 2 hrs and some point between 0 and 2 hrs	Flag justified = 4 Manual revision = 4 Extenuating circumstances = 1 Practice factors = 1 Decision rule = 1 Criteria revision = 6 Total flags reviewed = 17 1. GTT not reported, but dx and therapy indicate positive. *2-3. On a new diagnosis, a single fasting BS >200 should be acceptable. Two random BS >200 is acceptable; may not have weight loss.* **4. Rigid rule to document associated symptoms should not be necessary to establish diagnosis.** **5. Drop GTT requirement.** **6. Mirror ADA.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DIAGNOSIS ANALYSIS
Peer Reviewer Comments, Three State Combined Data
November 3, 1993

CONGESTIVE HEART FAILURE **CODE: 06**

Indicator	Criterion	Peer Review
Old Diagnosis Monitoring	Chest pain - once	Flag justified = 7 Manual revision = 5 Decision rule = 1 Total flags reviewed = 13 **1. Note indicates "review of systems done and was negative." This adequately documents that the patient was questioned as to difficulties related to CHF. The review of symptoms would include the specific criteria.**
Evidence for New Diagnosis	S3, pulmonary rales, raised JVP and/or pedal edema, enlarged heart OR left ventricular ejection fraction <40% OR signs of CHF on chest x-ray.	Flag justified = 11 Manual revision = 5 Criteria revision = 1 Total flags reviewed = 17 *1. All signs of CHF are not necessary; criteria need to allow for clinical variation.*
Search for Treatable Causes	HTN AND valvular heart disease AND AMI AND drug role AND anemia AND statement re thyroid disease	Flag justified = 6 Manual revision = 6 Criteria revision = 4 Total flags reviewed = 16 1. Echo showed septal akinesis; therefore, cause of CHF due to infarct - thyroid studies not needed. *2-4. Any one of six treatable causes should be acceptable.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa** (**bold**), and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

ANGIOTENSIN-CONVERTING ENZYME (ACE) INHIBITOR **CODE: 01**

Indicator	Criterion	Peer Review
Drug Indications	Score: Drug prescribed appropriately for: HTN, CHF, Bartter's syndrome	Flag justified = 3 Manual revision = 7 Extenuating circumstances = 1 Decision rule = 1 Criteria revision = 4 Total flags reviewed = 16 1. Although HTN not stated, Dr. clearly addressing elev. BP; criteria should allow other justification. 2. Criteria should expand time frame to look for diagnosis. 3. Criteria should allow LV dysfunction and CHF. 4. Need to expand crit. to allow HTN diag. where BP markedly elev./already on antihypertensives.
Drug-Condition Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Risk/benefit:** Angioedema, hyperkalemia, COPD, chronic cough	Flag justified = 4 Manual revision = 8 Decision rule = 1 Criteria revision = 5 Total flags reviewed = 18 1. COPD/asthma are not contraindications to the use of ACE inhibitors. *2. COPD not a contraindication to use of vasotec.* **3. Reviewer believes avoiding ACE inhibitors in COPD patients is a theoretical concern and not clinically justified.** **4. Capoten not contraindicated in patient with COPD.** **5. Delete COPD from list.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug-drug Contraindications		Flag justified = 10 Manual revision = 3 Extenuating circumstances = 2 Total flags reviewed = 15
Drug Monitoring	**Pre-existing Drugs:**	Flag justified = 7 Manual revision = 4 Patient factors = 1 Criteria revision = 6 Total flags reviewed = 18 **1. Old drug, no way of knowing when med started.**
	If renal dis., BUN or creat. - q. 3 mos	**2. Re-check of labs q 4-6 mos would be sufficient when in long-term use and patient stable.**
	BUN or creat. within 1 mo	*3. Lab testing interval may be too strict.* **4. All of the lab monitoring on starting ACE inhibitors would probably be revised to allow within 3 mos of drug initiation.** **5-6. Within 3 mos is more reasonable.**
	WBC and Diff. within 3 mos	*3. Lab testing interval may be too strict.* **4. All of the lab monitoring on starting ACE inhibitors would probably be revised to allow within 3 mos of drug initiation.**
	K+ within 1 mo	*3. Lab testing interval may be too strict.* **4. All of the lab monitoring on starting ACE inhibitors would probably be revised to allow within 3 mos of drug initiation.** **5-6. Within 3 mos is more reasonable.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (**bold**), and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

ANTIADRENERGIC - PERIPHERAL ALPHA BLOCKER **CODE: 03**

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: HTN, CHF, Raynaud's phen., bladder outlet obstr., benign prostatic hypertrophy	Flag justified = 2 Manual revision = 8 Criteria revision = 4 Total flags reviewed = 14 1. Question whether step therapy treatment mandatory for HTN. *2. Phrase "only if step 1 drug therapy failed" should be eliminated.* **3. Should be more flexibility in what first-line BP meds should be used first.** **4. Alpha-blockers should be acceptable first-line agents.**
Drup-Condition Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Risk/benefit:** Orthostatic hypotension	Flag justified = 1 Criteria revision = 1 Total flags reviewed = 2 **1. Physician is monitoring orthostatic BP. This is indicative that he/she aware of the risks of Hytrin.**
	Pre-existing Drugs: BP q 12 mos, standing	Flag justified = 8 Manual revision = 6 Practice factors = 1 Decision rule = 1 Criteria revision = 3 Total flags reviewed = 19 **1. Orthostatic BP not needed without symptoms.** **2. Reviewer does not believe that standing BP needs to be documented.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (**bold**), and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
	If angina, within 1 mo start of therapy educate patient re possible exacerbation	**1. PDR reviewed. No mention of angina or significant problem for use of Hytrin.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

ANTIADRENERGIC - BETA BLOCKER **CODE: 04**

Indicator	Criterion	Peer Review
Drug Indications	Score: Drug prescribed appropriately for: Angina, ventricular arrhyth., HTN, hypertrophic cardiomyopathy, myocardial reinfarct prophylaxis, pheochromocytoma, vascular headache prophylaxis, non-Parkinson's tremor, "stage-fright" anxiety, thyrotoxicosis, Barlow's synd., atrial arrhythmias	Flag justified = 2 Manual revision = 4 Practice factors = 1 Decision rule = 1 Criteria revision = 6 Total flags reviewed = 14 1. Appropriate to use B-blockers as first-step therapy for HTN. 2. Trandate OK as first-time antihypertensive. *3. "Only if other step 1 drugs have failed" should be eliminated. Stepped care for HTN has been replaced by an individualized approach.* *4. B-blockers are appropriate step drugs. No need to wait for failure of another drug to use B-blockers.* **5. Delete "Only if other step 1 drugs have failed."** **6. Add abnormal treadmill to criteria.**
Drug-Condition Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Absolute:** Heart block, asthma, COPD **Risk/benefit:** Mental depression, Raynaud's phen., intermittent claudication	Flag justified = 6 Manual revision = 7 Extenuating circumstances = 2 Criteria revision = 3 Total flags reviewed = 18 1. COPD should be relative contraindication. 2. COPD should be a relative, not an absolute contraindication. **3. COPD and asthma are relative contraindications and should be taken from absolute contraindication list.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (**bold**), and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug-Drug Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Absolute:** None **Risk/benefit:** Calcium channel blockers, oral antidiabetic agents, insulin, indomethacin, Timolol eyedrops	Flag justified = 11 Manual revision = 2 Criteria revision = 5 Total flags reviewed = 18 1. Lopressor and insulin started prior to study. 2. Drug initiated prior to study. *3. Frequently use calcium channel blockers with B-blockers in refractory HTN when no CHF.* **4. Consider removal of calcium channel blocker and indomethacin from list 2.** **5. B-blockers (including Inderal) are not contraindicated for all calcium channel blockers.**
Drug Monitoring	**Pre-existing drugs**	Flag justified = 11 Manual revision = 5 Patient factors = 1 Practice factors = 1 Criteria revision = 2 Total flags reviewed = 20
	BP - q 6 mos, regular	**1. With availability of home BP monitors and knowing patient reliability, it is not necessary to demand q 6 mos.**
	BP - q 12 mos, standing	**2. Reviewer disagrees with need to record and monitor a standing BP.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

ANTIADRENERGIC - CENTRALLY ACTING CODE: 05

Indicator	Criterion	Peer Review
Drug Indications	Score: Drug prescribed appropriately for: HTN; headache, vascular; nicotine dependence	Flag justified = 5 Manual revision = 6 Practice factors = 2 Decision rule = 1 Criteria revision = 3 Total flags reviewed = 17 1. Clonidine may be used for rapid treatment of accelerated HTN without step therapy. 2. Elevated BP and evidence of prior treatment with antihypertensives should be diagnostic of HTN. **3. We should allow alpha-blockers and central agonists to be first-line choices in certain situations.**
Drug-Condition Contraindications		Flag justified = 1 Manual revision = 1 Total flags reviewed = 2
Drug-Drug Contraindications		Flag justified = 1 Manual revision = 4 Total flags reviewed = 5
Drug Monitoring	**Pre-existing drugs**	Flag justified = 10 Manual revision = 2 Criteria revision = 6 Total flags reviewed = 18 **1. Only if orthostatic symptoms present, are supine and upright BP appropriate.** **2. Criteria for upright BP need to be changed.** **3. If orthostatic symptoms not present in dx, supine and upright BP are questionable.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
	Mental status (mood), once	**4. Reviewer disagrees that mood should be checked any more than other side effects that occur 1-3% of the time.** **5-6. Delete mood status check.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

CALCIUM CHANNEL BLOCKER CODE: 06

Indicator	Criterion	Peer Review
Drug Indications	Score: Drug prescribed appropriately for: Angina, HTN, Raynaud's phen., hypertrophic cardiomyopathy, migraine	Flag justified = 2 Manual revision = 7 Decision rule = 1 Criteria revision = 5 Total flags reviewed = 15 *1. Criteria should include CAD as well as angina.* **2. Using Calan to treat or prevent supraventricular arrhythmias also a reasonable indication.** **3. Add indications for tachyarrhythmias.** **4. Need to include indication: Arrhythmia, PSVT, and control of rate for chronic atrial flutter/fibrillation.** **5. Can include abnormal treadmill as criteria for cardizem use.**
Drug-Condition Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Absolute:** Heart block, sick sinus synd., Wolff-Parkinson-White synd. **Risk/benefit:** Bradycardia, CHF	Manual revision = 8 Criteria revision = 4 Total flags reviewed = 12 **1. Old drug-old diagnosis. Cardizem is to be used with caution in patients with impaired ventricular function. Have to assume dr. knows something and does not have to justify in writing every time a decision is made.** **2. Patient had permanent pacemaker and, therefore, was protected from the potential complications of this med. Revise list to state heart block, sick sinus syndrome unless pacemaker.** **3. Procardia can be used in CHF unless CHF very severe or worsened with procardia.** **4. Calan can be used in CHF as long as patient is watched and no signs of worsening CHF.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug Monitoring		Flag justified = 4 Manual revision = 3 Decision rule = 1 Criteria revision = 2 Total flags reviewed = 10
	BP q 12 mos, standing	**1. If BP stable and no other stated symptoms, then need not be standing.**
	If taking beta blocker, pulse rate and rhythm or ECG within 1 mo start of therapy	**2. If pulse rate OK, then ECG is not needed.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (bold), and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

DIURETIC - LOOP CODE: 08

Indicator	Criterion	Peer Review
Drug Indications	Score: Drug prescribed appropriately for: Edema/pulm. edema, HTN, hypercalcemia	Flag justified = 3 Manual revision = 1 Criteria revision = 5 Total flags reviewed = 9 1. Criteria need to be expanded to allow for old diagnosis at first visit of patient on antihypertensive medication indicating was HTN. 2. Appropriate to give bumex without CHF and current congestion in lungs by CXR. Difficult to differentiate from edema in some cases. 3. CHF appropriate indication for lasix. **4-5. CHF should be an appropriate diagnosis for diuretic use.**
Drug-Condition Contraindications	Score: No contraindications or use of drug justified despite contraindications: Risk/benefit: BUN >50 or creat. >2.5	Flag justified = 3 Manual revision = 4 Extenuating circumstances = 3 Criteria revision = 2 Total flags reviewed = 12 **1. Not a concern with patient having terminal metastatic cancer.** **2. Not necessary to make statement of risk/benefit when at times this may be the drug of choice with BUN >50 and/or creatinine >2.5**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Indicator	Criterion	Peer Review
Drug-Drug Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Absolute:** Lithium **Risk/benefit:** Long-term high-dose corticosteroid	Flag justified = 2 Manual revision = 4 Extenuating circumstances = 1 Criteria revision = 3 Total flags reviewed = 10 *1. It is not contraindicated to use steroids and diuretics when clinically necessary.* **2. Patient on long-term care so risk/benefit may have been done prior to time frame of study.** **3. Change definition of long-term therapy.** **4. Reviewer can find nothing in PDR that addresses risk of using furosemide and prednisone. The real concern is: "Does the dr. address, in print, the risk/benefit ratio?" Reviewer would like to discuss this further.**
Drug Monitoring	K+ q 6 mos	Flag justified = 11 Manual revision = 1 Patient factors = 2 Criteria revision = 1 Total flags reviewed = 15 1. *Should accept any time in 6 mos; q 6 mos is too rigid.*
	Na+ q 6 mos	1. *Should accept any time in 6 mos; q 6 mos is too rigid.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa** (bold), and *Maryland* (*italics*).

DRUG ANALYSIS

Peer Reviewer Comments: Three State Combined Data

November 3, 1993

POTASSIUM-SPARING DIURETIC CODE: 09

Indicator	Criterion	Peer Review
Drug Indications		Flag justified = 1 Total flags reviewed = 1
Drug-Drug Contraindications	Score: No contraindications OR use of drug justified despite contraindications: **Absolute:** Oral potassium supplements, ace inhibitors **Risk/benefit:** Aldactone unaccompanied by potassium-losing diuretic except in hepatic failure	Criteria revision = 1 Total flags reviewed = 1 1. Patient on spironolactone/ACE inhibitor prior to beginning of study, so explanation would not be available. K+ checked frequently to avoid risk of hyperkalemia.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

DIURETIC - THIAZIDE-TYPE **CODE: 10**

Indicator	Criterion	Peer Review
Drug Indications	Score: Drug prescribed appropriately for: Edema/pulm. edema, HTN, diabetes insipidus, renal calculi	Flag justified = 2 Manual revision = 4 Extenuating circumstances = 1 Criteria revision = 1 Total flags reviewed = 8 **1. Should add CHF to dx list here.**
Drug-Condition Contraindications		Flag justified = 13 Manual revision = 3 Total flags reviewed = 16
Drug-Drug Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Absolute:** None **Risk/benefit:** Oral antidiab. agents, insulin, long-term high-dose corticosteroid therapy	Flag justified = 3 Manual revision = 8 Patient factors = 1 Criteria revision = 3 Total flags reviewed = 15 **1. If FBS, HbA1C, 2 hr pp BS are within guidelines, then no physician review needed.** **2. If BS controlled, then should not flag case.** **3. Insulin should be a risk/benefit contraindication.**
Drug Monitoring	**Pre-existing Drugs**	Flag justified = 7 Patient factors = 1 Criteria revision = 10 Total flags reviewed = 18 1. If combination antihypertensive is changed but diuretic portion remains the same, diuretic criteria for old drug should be used.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

Indicator	Criterion	Peer Review
	BUN or creat. q 6 mos	**2a. Consider less rigid criteria; patient conditions are so variable.** **3-4. In a stable patient, monitor BUN/creat., sodium, potassium annually.** **5a. Change monitoring to once per yr.** *6-7. Within 1 mo start of therapy and then, if within normal range, q 3 to 6 mos thereafter.*
	Na+ q 6 mos	**2b. Consider less rigid criteria; patient conditions are so variable.** **3-4a. In a stable patient, monitor BUN/creat., sodium, potassium annually.** **5b. Delete requirement.** **8. Reviewer would not routinely check a stable patient q 6 mos; would suggest annually.** *9. Criteria should say within 6 mos or 2x per year.*
Drug Monitoring (cont)	K+ q 6 mos	**2c. Consider less rigid criteria; patient conditions are so variable.** **3-4b. In a stable patient, monitor BUN/creat., sodium, potassium annually.** **5c. Delete requirement.**
	Na+ within 1 mo	*10. Should be acceptable to have within 3 mos start of therapy.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (**bold**), and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

ACE INHIBITOR + THIAZIDE DIURETIC CODE: 11

Indicator	Criterion	Peer Review
Dose and Frequency	Score: New drug prescribed appropriately for: HTN if stepped therapy fails	Criteria revision = 1 Total flags reviewed = 1 1. CHF is appropriate indication for both ACE inhibitor-capoten and diuretic HCTC and thus for capozide.
Drug-Condition Contraindications	Score: No contraindications, OR use of drug justified despite contraindications: **Absolute:** Several renal impairment, hypercalcemia, hyperuricemia **Risk/benefit:** Angioedema, hyperkalemia, gout, COPD, impaired hepatic function, SLE	Criteria revision = 1 Total flags reviewed = 1 1. Patient on capozide prior to study; justified likely when drug began.
Drug Monitoring	If diabetes, blood glucose q 3 mos	Flag justified = 1 Manual revision = 1 Criteria revision = 1 Total flags reviewed = 3 1. BS ok q 4 mos in stable patient.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

BETA BLOCKER + THIAZIDE CODE: 13

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: HTN if stepped therapy fails	Criteria revision = 1 Total flags reviewed = 1 1. Limited time scope of study did not allow for diagnosis of HTN.
Drug-Condition Contraindications	Score: No contraindications OR use of drug justified despite contraindications: **Absolute:** Severe renal impairment, hypercalcemia, CHF, heart block, asthma, emphysema or non-allergic bronchitis **Risk/benefit:** Mental depression, Raynaud's intermittent claudication, gout, impaired hepatic function, SLE	Manual revision = 1 Criteria revision = 1 Total flags reviewed = 2 1. COPD is a relative contraindication for beta-blocker.
Drug Monitoring		Flag justified = 3 Total flags reviewed = 3

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

VASODILATOR AND THIAZIDE DIURETIC CODE: 17

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: HTN if stepped therapy fails	**1. Patient apparently on Ser-Ap-Es prior to apresoline-esidrix; unable to determine extent of prior treatment.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS

Peer Reviewer Comments: Three State Combined Data

November 3, 1993

COMBINATION: DIURETIC (THIAZIDE-TYPE)+POTASSIUM-SPARING-TYPE CODE: 18

Indicator	Criterion	Peer Review
Drug Indications	Score: Drug prescribed appropriately for: Edema/pulm. edema, HTN, diabetes insipidus, renal calculi, CHF	Flag justified = 7 Manual revision = 7 Extenuating circumstances = 1 Criteria revision = 1 Total flags reviewed = 16 1. Criteria should be expanded to include prior treatment of HTN as diagnostic. Elevated BP should also equal HTN in patient with probable old treatment.
Drug-Condition Contraindications		Flag justified = 8 Manual revision = 4 Total flags reviewed = 12
Drug-Drug Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Absolute:** Oral potassium supp., ACE inh. **Risk/benefit:** Oral antidiab. agents, insulin, long-term high-dose corticosteroid therapy	Flag justified = 9 Manual revision = 5 Extenuating circumstances = 1 Criteria revision = 3 Total flags reviewed = 18 1. Hypokalemia is an adequate justification for KCl. **2. Oral hypoglycemics not a major problem. Awareness of K+ and renal functions far more important.** **3. ACE inhibitors are commonly given with dyazide.**
Drug Monitoring	BP q 6 mos, regular	Flag justified = 9 Manual revision = 3 Patient factors = 1 Extenuating circumstances = 2 Criteria revision = 4 Total flags reviewed = 19 *1. A suggestion for all time frames is that 6 mos ±4 wks would be better than ±2 wks.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (bold), and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
	BP q 12 mos, standing	2. *Standing BPs are not routinely done.* **3. Standing BP probably not necessary unless there are complaints of orthostatic hypertension.**
	BUN or creat. q 6 mos	**4. Should not be so rigid. Criteria call for 6 mos monitoring; patient in office at 7 mos.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

NITRATE CODE: 21

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Angina, CHF	Flag justified = 7 Manual revision = 2 Criteria revision = 7 Total flags reviewed = 16 1. See IHD flag - incorrect - Isordil was indicated due to strong suspicion of IHD. 2. CAD appropriate indication for NTG 3. Synonym list too restrictive. ISHD listed as diagnosis for ischemic heart disease - appropriate indication for nitro-dur. 4. History of coronary artery bypass should be allowed as criteria for diagnosis of CAD. 5. Isordil indicated due to strong suspicion of IHD. 6. *Dx of angina should include other symptoms such as chest pain, r/o angina, and empiric trials of prn meds.* **7. Add CAD to list.**
Drug-Condition Contraindications		Flag justified = 2 Manual revision = 2 Patient factors = 1 Total flags reviewed = 5

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug-Drug Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Transdermal nitrates not to be used continuously for 24 hrs.	Flag justified = 3 Manual revision = 5 Extenuating circumstances = 2 Practice factors = 1 Criteria revision = 2 Total flags reviewed = 13 *1. Patient should be able to use sublingual nitro on prn basis for acute chest pain, even if already on nitro patch.* **2. Continuous use of transdermal nitrates is controversial. Until conclusive data is available, the standard of care must allow for individual physician judgment.**
Drug Monitoring	If angina, educate patient re possible exacerbation. If CHF, wt and lung exam within 1 mo start of therapy	Flag justified = 8 Manual revision = 5 Extenuating circumstances = 1 Practice factors = 1 Criteria revision = 3 Total flags reviewed = 18 **1. PDR does not list this criteria.** **2. Isorbide seldom worsens angina, so do not need to ask patient about angina.** **3. Nitrostat rarely worsens angina, so do not need to require evidence that physician asked.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

INSULIN — CODE: 31

Indicator	Criterion	Peer Review
Drug-Drug Contraindications	Score: No contraindications OR drug justified despite contraindications: **Risk/benefit:** Fenfluramin, long-term high-dose corticosteroid therapy, beta blockers	Flag justified = 4 Manual revision = 3 Extenuating circumstances = 1 Criteria revision = 2 Total flags reviewed = 10 1. Insulin/lopressor started prior to study. **2. Cannot find a contraindication for fenfluromine and insulin.**
Drug Monitoring	Blood glucose q 3 mos	Flag justified = 8 Manual revision = 2 Practice factors = 1 Patient factors = 1 Extenuating circumstances = 1 Criteria revision = 5 Total flags reviewed = 18 1. In a stable patient with Accuchek at home, follow-up office BS needs to be done only q 3-6 mos. **2. A 3-mo comment on BS not necessary when frequent mention of home BS monitoring.**
	HgbA1C or fructosamine - once	*3. This criteria is inappropriate where the time frame does not fit the circumstances or where blood sugars reflect the patient's status adequately.* *4. This should be contingent upon random blood sugar findings.* **5. Have reservation about requiring HgbA1C due to cost. Not clear that therapy would be changed by results.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

ORAL HYPOGLYCEMIC CODE: 32

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Diabetes mellitus, non-insulin dependent or diabetes insipidus, partial central	Flag justified = 3 Manual revision = 3 Practice factors = 1 Criteria revision = 1 Total flags reviewed = 8 1. Evidence of positive GTT indicates DM and makes treatment with diabeta appropriate.
Drug-Condition Contraindications	Score: No contraindications or use of drug justified despite contraindications: **Risk/Benefit:** patient > 65 (chlorpropamide), renal impairment (glyburide)	Flag justified = 4 Decision rule = 2 Criteria revision = 2 Total flags reviewed = 8 1. Study period does not include time frame when drug was started. Risk/benefit should have been done then. **2. Records prior to 1989 are required to adequately review patient care. Patient has long-standing DM and has been well controlled on chlorpropamide. No need to change med or dose simply due to age, with no evidence of lower BS and hypoglycemia.**
Drug-Drug Contraindication	Score: No contraindications or use of drug justified despite contraindications: **Absolute:** beta blockers, insulin **Risk/benefit:** Phenytoin, long-term high-dose corticosteroids, anticoagulants, thiazides, calcium channel blockers, long-term high-dose salicylates, captopril	Flag justified = 5 Manual revision = 6 Extenuating circumstances = 1 Decision rule = 1 Criteria revision = 5 Total flags reviewed = 18 1. It is not clear that the insulin and the diabeta are contraindicated together. 2. Patient placed on diabeta/inderal prior to beginning of study. Risk/benefit would have been done at start of medication. 3. *Beta blockers, insulin are relative contraindications.* **4. Such a rare instance not necessary to document risk/benefit every time the two meds are given simultaneously.** **5. No evidence that use of oral hypoglycemics with insulin is contraindicated.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug Monitoring	Blood glucose: q 3 mos	Flag justified = 9 Manual revision = 5 Patient factors = 1 Criteria revision = 4 Total flags reviewed = 19 1. In stable NID DM non-insulin requiring BS checks q 3-6 mos appropriate and glyHgb on HgbA1C not necessary. **2. May be appropriate to monitor blood glucose q 6 mos instead of requiring q 3 mos.**
	HgbA1C: once	3. In stable NID DM non-insulin requiring BS checks q 3-6 mos appropriate and glyHgb on HgbA1C not necessary. *4. Do not need HgbA1C if diabetes well-followed by blood sugar.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (**bold**), and *Maryland* (*italics*).

CORTICOSTEROID CODE: 41

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Anti-inflamm., anti-supp., severe asthma, acute/chron. lymphocytic leukemia, non-Hodgkin's / Hodgkin's lymphomas, multiple myelomas, metastatic breast cancer	Flag justified = 6 Manual revision = 8 Criteria revision = 4 Total flags reviewed = 18 **1. Reviewer would add COPD to indications for prednisone.** **2. Include steroid use to minimize post-herpetic neuralgia (in situations where cost of zovirax is prohibitive).** **3. Need to add COPD, chronic.** **4. Add temporal arteritis.**
Drug-Condition Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Untreated TB **Risk/benefit:** Treated TB, diabetes, myocardial infarct preceding mo, esophagitis, gastritis, peptic ulcer	Flag justified = 7 Manual revision = 8 Extenuating circumstances = 2 Total flags reviewed = 17 **1. Physician aware of prednisone side effects, but patient had steroid-dependent disease.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa** (bold), and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug-Drug Contraindications	No contraindications OR drug justified despite contraindications: **Absolute:** Aspirin (high-dose long-term use), nonsteroidal anti-inflamm. drugs (long-term) **Risk/benefit:** Oral antidiabetic agents, thiazides, loop diuretics	Flag justified = 3 Manual revision = 3 Patient factors = 1 Decision rule = 1 Criteria revision = 10 Total flags reviewed = 18 1. Combination of corticosteroids and NSAIDs should be in the relative CI category. *2-3. Prednisone + diuretics used frequently with no significant clinical problems.* *4. Why not Lasix + prednisone?* *5. Steroid creams are not contraindicated with thiazide diuretics.* *6. Should differentiate between short course and chronic long-term use.* **7. An explanation of risk/benefit should not be necessary in outpatient when physician is summarizing meds and care started in preceding hospitalization. Such an explanation would have been done in the hospital**
		8-9. Prednisone and loop diuretic is theoretical concern only and generally a safe combination. Concerned that requiring too much documentation. **10. Aspirin and NSAIDs are relative contraindications. Also, not necessary to document thiazides and loop diuretics.**
Drug Monitoring	Eye exam once	Flag justified = 6 Manual revision = 1 Practice factors = 1 Criteria revision = 9 Total flags reviewed = 17 **1. Ophthalmic exams frequently not recorded in first physician chart.**
	If HTN, BP q 3 mos, with BP taken standing once/yr.	**2. If orthostatic symptoms present, BP supine and upright indicated; otherwise, not required.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
	PPD before start of drug	3-7. PPD is indicated for long-term corticosteroid usage, not short-term use. **8-9. When physician knows from the start that prednisone will be for a short time, then TB monitoring is not necessary.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS

Peer Reviewer Comments: Three State Combined Data

November 3, 1993

Nonsteroidal Anti-inflammatory Analgesic (Nonsalicylate) (NSAID) — Code: 42

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Arthritis, ankylosing spondylitis, Reiter's dis., pain, gout, inflamm. (non-rheumatic), headache (vasc. or migraine), Bartter's synd., phenylbutazone, indomethacin.	Flag justified = 3 Manual revision = 12 Extenuating circumstances = 1 Decision rule = 1 Criteria revision = 3 Total flags reviewed = 20 1. Disagree that indocin cannot be used as first-time therapy. 2. Obvious joint deformity implies arthritis. **3. Any NSAID would be appropriate for acute gout.**
Drug-Condition Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Aspirin allergy, anemia, renal impairment, PVD, bleeding diseases, CHF, jaundice, upper GI disease, asthma **Risk/benefit:** OA, RA - indomethacin	Flag justified = 5 Manual revision = 4 Patient factors = 1 Decision rule = 1 Criteria revision = 6 Total flags reviewed = 17 1. Type of anemia needs to be specified here. Anemia due to acute bleeding would contraindicate use of feldene. 2. CHF is not an absolute contraindication to naprosyn. 3. *Nalton not contraindicated with controlled CHF.* **4. Contraindication should be listed as relative rather than absolute.** **5. Need new criteria; eliminate list 2.** **6. No information available that prior diagnosis was CHF.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug-Drug Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Aspirin, anticoagulants **Risk/benefit:** Anticoagulants, methotrexate, chron. alcohol	Flag justified = 8 Manual revision = 2 Criteria revision = 8 Total flags reviewed = 18 1. Patient on very low dose ASA for cardiovascular prophylaxis. MD protecting patient GI mucosa with histamine 2 blocker. 2. MD documents that patient was warned. *3. Patient on low-dose aspirin. This is not a contraindication.* **4. Current prescription 1 aspirin/day with other NSAID not absolute contraindication.** **5-8. Low-dose ASA and NSAID is OK.**
Drug Monitoring		Flag justified = 11 Extenuating circumstances = 3 Criteria revision = 4 Total flags reviewed = 18 **1. OTC drugs cannot be monitored by physician on regular basis.** **2. Old drug criteria not justified. Patient should have yearly follow-ups of Hgb, Hct, BUN, creat., SGOT, SGPT.**
	BUN or creat q 6 mos	**3. Yearly**
	Urinalysis q 6 mos	**3. Yearly** **4. Not really necessary to monitor NSAID therapy**

Inter-rater scores based on three-state data.

Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

SALICYLATE CODE: 43

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Chron. pain, nonrheumatic inflamm., rheumatic arthritis, osteoarthritis, migraine prophylaxis	Flag justified = 1 Manual revision = 2 Decision rule = 1 Criteria revision = 14 Total flags reviewed = 18 1-2. Expand criteria to include cardiovascular prophylaxis as indication for aspirin. 3. Aspirin appropriately used for cardiovascular prophylaxis. 4. Flu syndrome should be considered acute inflamm. *5-6. Add platelet aggregation inhibition to criteria.* *7. Expand to include prophylaxis for arteriosclerotic disease.* *8. Add CAD and CVD as indicators.* *9. Add prevention of MI and CVA to indications.* **10. Two diagnoses are present; positive degenerative disease and prevention of phlebitis in chart; add to criteria.** **11. ASA for prevention in setting of angina/MIS and in setting of TIAs should be OK.** **12-14. Need to add MI prevention to criteria.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug-Condition Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Gastrointestinal bleeding/ulcers, bleeding disorders **Risk/benefit:** Nasal polyps, adult-onset asthma, chron. alcohol use	Flag justified = 9 Manual revision = 6 Practice factors = 1 Criteria revision = 1 Total flags reviewed = 17 **1. Result of UGI not provided in chart. Must assume from therapy that UGI negative.**
Drug-Drug Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Allergy to aspirin or other NSAIDs, anticoagulants, NSAID **Risk/benefit:** Oral antidiabetic agents, furosemide, methotrexate	Flag justified = 5 Manual revision = 4 Criteria revision = 9 Total flags reviewed = 18 1. Lasix/furosemide will not cause significant rise in salicylate level in patients on low-dose aspirin. *2. Common to use ASA at low dose with NSAIDs.* *3. ASA + indocin poses no real problem.* **4. Low-dose aspirin use with NSAIDs should be under a relative risk setting, not absolute contraindication.** **5-7. ASA in low dose with NSAID OK.** **8. One ASA/day given for blood thinner does not require justification.** **9. AMA Drug Evaluations 1992, p. 956 Interactions: notes large doses of salicylates may increase risk of hypoglycemia. Criterion lacks specifics.**
Drug Monitoring		Flag justified = 4 Manual revision = 4 Criteria revision = 10 Total flags reviewed = 18 1. Patient on aspirin for cardiovascular prophylaxis - very low dose? Need criteria. (Should they be different for low-dose aspirin?)

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
	Hx for GI distress q 6 mos	**2. Reviewer thinks any chart mention of GI distress without having to discuss possible role of ASA.** 3. For patient taking aspirin for cardiovascular prophylaxis, probably would be appropriate to inform patient of potential for GI distress and check Hct annually. 4. Criteria are not applicable to very low-dose aspirin used prophylactically. Hct and GI assessment not necessary. **5. Patient on ASA for coronary artery disease. Not necessary to ask about gastric distress or have Hct checked q 6 mos.** **6. Suggest annual check for GI distress.**
	HCT q 6 mos	7. Patient taking low-dose aspirin, Hct q yr adequate. *8. Hct within 6 mos a better criterion.* **9. Six mos too rigid; suggest annually.**
	Hx for GI distress within first 6 mos	*10. Monitoring not necessary for low-dose ASA therapy.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (bold), and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

BETA-ADRENERGIC AGONIST (ADRENERGIC BRONCHODILATOR) - INHALED **CODE: 51**

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Bronch. asthma, bronchiectasis, chron. bronch., and emphysema	Flag justified = 2 Manual revision = 5 Criteria revision = 10 Total flags reviewed = 17 1. Criteria needs to be expanded - include obstructive lung disease (esp. when PFTs indicate such). 2. Bronchodilators are appropriate therapy in acute and chronic bronchitis. 3. Appropriate to treat symptomatic bronchitis, acute or chronic, with proventil. 4. Any condition that causes bronchospasm/wheezing is appropriate indication for proventil. *5. Criteria should be broadened to include wheezing.* *6. Criteria should be broadened to include associated symptoms.* *7. Criteria should include persistent bronchospasm, i.e., wheezing or non-productive, dry persistent cough.* **8. Add to criteria, #4 acute bronchitis with wheeze (e.g., bronchospasm).** **9. Patient had COPD as a diagnosis. For clarity, revise #3 to: "chronic bronchitis, emphysema, and/or COPD.** **10. COPD, emphysema, asthma, chronic bronchitis are terms often used interchangeably; while they are different disease processes, all are appropriate for beta agonist.**
Drug-Condition Contraindications		Flag justified = 1 Manual revision = 2 Total flags reviewed = 3
Drug-Drug Contraindications		Manual revision = 2 Practice factors = 1 Total flags reviewed = 3

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (bold), and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug Monitoring	Lung exam q 6 mos	Flag justified = 8 Manual revision = 2 Extenuating circumstances = 2 Patient factors = 1 Criteria revision = 5 Total flags reviewed = 18 1. Monitoring time frame for checking pulm. exam on new albuterol should be more liberal. **2. With mild, non-hypoxemic bronchitis, ok to extend following interval.** *3. Time frame should be more liberal for maintenance drug for chronic condition.* *4. Should allow greater flexibility in monitoring according to patient's clinical status and symptoms.*
	Lung exam within 1 mo start of therapy	5. Due to relative lack of toxicity of ventolin, doubt need to have lung exam in 1 mo, esp. when using PRN appropriate to evaluate.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

XANTHINE-DERIVATIVE - ORAL **CODE: 53**

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Bronch. asthma, pulmonary emphysema, chron. bronchitis	Flag justified = 5 Manual revision = 1 Decision rule = 1 Criteria revision = 5 Total flags reviewed = 12 1. Criteria do not allow for determination of diagnosis due to limited time span. Slo-bid may have been an old medication. 2. Theophylline appropriate for chron. bronchitis. 3. Theo-dur indicated for any type of bronchospasm except that due to foreign body obstruction. 4. Theophyline appropriate for any acute type of bronchospasm. **5. Reviewer should have been able to accept "probable obstructive lung disease." This was listed in chart.**
Drug-Condition Contraindications	Score: No contraindications OR drug justified despite contraindications: Risk/benefit: Gastritis or peptic ulcer, CHF, liver dis., prostatic hypertrophy, alcoholism	Flag justified = 9 Manual revision = 3 Decision rule = 2 Criteria revision = 4 Total flags reviewed = 18 *1. No clinical reason not to use theodur in patients with CHF.* *2. Broaden criteria to use with specific types of CHF.* **3. Alcoholism true only with liver disease. Eliminate prostate hypertrophy.** **4. Eliminate CHF.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug-Drug Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** On other xanthine derivative **Risk/benefit:** Cimetidine, erythromycin, beta blockers, furosemide, propanolol	Flag justified = 7 Manual revision = 3 Criteria revision = 7 Total flags reviewed = 17 1. Patient placed on very small dose of theo- doubt level would be elevated with tagamet. *2. Question adverse relationship between xanthine and beta blockers.* *3-4. No problem using theodur + lasix.* **5. Reviewer does not feel that lasix use with theo-dur should be a contraindication.** **6. Reviewer questions furosemide listed in risk/benefit.** **7. Old drugs; cannot justify whether risk/benefit was done.**
Drug Monitoring	Serum theophylline q 3 mos	Flag justified = 8 Manual revision = 3 Criteria revision = 6 Total flags reviewed = 17 1-2. Appropriate to check theo level within mo after starting medication. In a stable patient with no clinical signs of toxicity, theo level q 6 mo - 1 yr would be appropriate. **3-5. Theo level check required q 3 mos in stable patient is too often.** **6. Criteria too strict and impractical.**
	Serum theophylline within 1 wk start of therapy	*5. One-wk check is too strict.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

BETA-ADRENERGIC AGONIST + XANTHINE-DERIVATIVE (ORAL) CODE: 54

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Bronch. asthma in patients who cannot use metered inhaler correctly and not in patients >65.	Criteria revision = 2 Total flags reviewed = 2 1. Guaifed is a decongestant, not a xanthine - used properly. 2. Guaifed is not a xanthine.
Drug-Condition Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Angina, hyperthyroidism, cardiac arrhythmias, HTN, gastritis or peptic ulcer, active or h/o CHF	Criteria revision = 2 Total flags reviewed = 2 1. Guaifed is a decongestant. Contraindications do not apply. 2. Guaifed is not a xanthine.
Drug Monitoring		Flag justified = 2 Criteria revision = 1 Total flags reviewed = 3 1. Guaifed is a decongestant; monitoring actions do not apply.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

BETA-ADRENERGIC AGONIST (ORAL) CODE: 55

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Bronch. asthma in patients who cannot use metered inhaler correctly, bronchiectasis, ephedrine not in patients >65, isoproterenol sublingual not in patients >65	Criteria revision = 3 Total flags reviewed = 3 1. COPD is appropriate indication for proventil. 2. Appropriate to use brethine for bronchospasm. 3. Wheezing is appropriate indication for brethine.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

CEPHALOSPORIN | CODE: 61

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: ***First Generation:** Lower resp. tract infec., urinary tract infec., skin/soft tissue infec., bone infec. **Second Generation:** *Lower resp. tract infec. *Urinary tract infec. *Skin/soft tissue infec. *Bone infec. Otitis media, urinary tract infec., gonorrhea-cefuroxime **Third Generation:** Urinary tract infec.	Flag justified = 5 Extenuating circumstances = 1 Criteria revision = 10 Total flags reviewed = 16 1. URI and pharyngitis are appropriate indications for Keflex. 2. Cellulitis is an appropriate indication for Duricef. 3. Keflex is appropriate first line for skin infections. 4. Ceftin appropriate first line for URI. 5. Keflex is appropriate first line therapy for bronchitis. 6. Ceclor is an appropriate therapy for pharyngitis. **7. Criterion too restrictive. Velosef is an appropriate choice of drug.** **8. Cephalosporins, first generation - add indication acute diverticulitis. Also add primary indications for cefaclor and ceftin for acute or chronic sinusitis without requiring penicillin/sulfa history of failure or allergy.** **9. Appropriate use of cephluxin for skin and soft-tissue infection. Erythromycin would have been only other choice.** **10. First generation cephalosporins should be included in first-line drugs (with penicillin, E'mycins, sulfas, etc.).**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug-Condition Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Previous allergic reaction to cephalosporin **Risk/benefit:** Colitis, penicillin allergy, BUN >50 or creat. >2.5	Flag justified = 5 Manual revision = 4 Patient factors = 1 Criteria revision = 8 Total flags reviewed = 18 1. Question whether penicillin allergy is a contraindication to cephalosporin. 2. Use of cephalosporin with penicillin allergy does not need additional assessment. *3-4. Cephalosporins not usually a problem in someone with a "soft" history of penicillin allergy.* **5. Penicillin only if allergy is anaphylactic in nature.** **6. No true correlation between penicillin allergy and cephalosporin allergy.** **7. Reviewer would allow cephalosporins in patients with penicillin allergy unless more serious or anaphylactic reaction.** **8. Reviewer suspects ceclor is used frequently with penicillin allergy being noted in chart. Does not think necessary to document.**

* Only if organism resistant or unresponsive to other anti-infectives or patient allergic to penicillins or sulfonamides or trimethoprim-sulfamethoxazole.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

ERYTHROMYCIN CODE: 62

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: **If allergic to penicillin:** Strep. pneumonia, strep. tonsillitis, community-acquired bronch., prophylaxis for dental procedures, otitis media, skin/soft tissue infec. Campylobacter infec., pertussis treatment or prophylaxis	Flag justified = 5 Criteria revision = 13 Total flags reviewed = 18 1. Erythromycin is appropriate therapy for URI. 2. Pharyngitis is appropriate indication for erythromycin. 3. URI is appropriate indication for erythromycin. 4. Erythromycin is appropriate for uncomplicated skin infec. 5. Erythromycin is appropriate for bronchitis, pleurisy, and URI. 6. Bronchopneumonia is appropriate indication for erythromycin. *7. Erythromycin can be used as a primary drug for bronchitis or other gram positive infection of respiratory tract.* **8. Should list possible atypical infections (e.g., mycoplasma, Legionnaires' dis., chlamydia) as first indications for erythromycin use.** **9. Erythromycin products are not restricted to penicillin-allergic patients.** **10. Add a #9 respiratory infection to criteria.** **11. Erythromycin is justified for sinusitis (even if PCN allergy not documented). Seems other antibiotic option would be more appropriate since patient already having gastritis.** **12. The clinical criteria need to be rewritten (see Guide to Antimicrobial Therapy 1992). No. 1 criteria inappropriately grouped. Strep tonsillitis, otitis media, and pertussis would be very rare in elderly patients. The Rx on this patient is appropriate and fits skin and soft tissue infections.** **13. Person has URI. Antibiotics can be justified due to tendency of this infection to descend to chest.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
	If treatment prolonged >2 wks and if on anticoagulants, prothrombin time within 1 wk start of therapy	Manual revision = 2 Criteria revision = 1 Total flags reviewed = 3 **1. Short-term erythromycin with appropriate monitors of prothrombin.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa** (bold), and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

PENICILLIN **CODE: 63**

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: **PEN V only** - Infec. of skin and soft tissue, pharynx infec., streptococcal or pneumococcal lower respiratory infec., infec. due to non-resistant streptococci and staphylococci, prophylaxis for dental proc.	Flag justified = 1 Criteria revision = 1 Total flags reviewed = 2 1. PCN appropriate treatment for sinusitis.
Drug-Condition Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Previous allergic reaction to penicillin or cephalosporin **Risk/benefit:** GI disease	Flag justified = 1 Criteria revision = 1 Total flags reviewed = 2 1. History of GI disease is not a contraindication to use of PCN.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

TETRACYCLINE CODE: 64

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Acute/chron. bronchitis, pneumonia, pelvic inflamm. dis., urinary tract infec., chron. prostatitis, Lyme dis., chlamydia, epididymo-orchitis	Flag justified = 8 Manual revision = 1 Criteria revision = 9 Total flags reviewed = 18 1. Tetracycline is appropriate therapy for URI. 2. Doryx is appropriate therapy for URI. 3. Tetracycline ok for labyrinthitis/vertigo. 4. Tetracycline ok for URI. *5. List should include rhinophyma.* *6. List needs to be more extensive.* *7. Add sinusitis.* **8-9. TCN useful in chronic or acute sinusitis.**
Condition Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** BUN >50 or creat. >2.5.	Manual revision = 1 Criteria revision = 2 Total flags reviewed = 3 1. No contraindicated condition. **2. PDR lists only absolute contraindication is an allergy to tetracycline.**
Drug Contraindications	Score: No contraindications OR drug justified despite contraindications: **Risk/benefit:** Anticoagulants	Criteria revision = 2 Total flags reviewed = 2 **1. Cannot find documentation of tetracycline and coumadin interaction (potential).** **2. Reviewer not aware of a risk of using tetracycline and an anticoagulant simultaneously.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
Drug Monitoring	**Newly Prescribed Drugs:**	Flag justified = 3 Manual revision = 2 Decision rule = 1 Criteria revision = 1 Total flags reviewed = 7
	If minocycline, warn re vestibular effect.	**1. Rx phoned in - patient not there to tell monitors.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

AMINOPENICILLINS CODE: 65

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Otitis media, sinusitis, shigella gastroenteritis, urinary tract infec., acute prostatitis, pelvic inflamm. dis., acute epididymo-orchitis, urethral gonorrhea, pneumonia	Flag justified = 4 Extenuating circumstances = 1 Criteria revision = 13 Total flags reviewed = 18 1-2. Bronchitis appropriate indicator for augmentin. 3. Augmentin appropriate for soft tissue infec. 4. Ampicillin appropriate for bronchitis. 5. Ampicillin appropriate for otitis/pharyngitis. 6. Augmentin appropriate for URI. *7. List should include more dx and symptoms in lieu of dx.* *8. Severe COPD with exacerbation could warrant Rx.* *9. Include bronchitis.* *10. Giving antibiotics should be justified in record up to 1 wk prior to prescribing.* **11. Would find it acceptable to Rx pharyngitis/tonsillitis.** **12. URI could be listed as a criterion for the use of amoxil.** **13. Recommended that a frequent use of antibiotic (amoxicillin) be allowed on indications lists as Rx for URI.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa** (**bold**), and *Maryland* (*italics*).

DRUG ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

SULFONAMIDE + TRIMETHOPRIM CODE: 72

Indicator	Criterion	Peer Review
Drug Indications	Score: New drug prescribed appropriately for: Lower urinary tract infec., pyelonephritis, prophylaxis of urinary infec., sinusitis, chron. bronchitis, chron. prostatitis, shigellosis, otitis media, traveler's diarrhea, salmonellosis	Flag justified = 7 Manual revision = 3 Decision rule = 1 Criteria revision = 7 Total flags reviewed = 18 1. Bactrim appropriate for acute prostatitis. 2. Septra appropriate for bronchitis. *3. Include symptoms of UTI.* **4. Revise #6 for "acute or chronic prostatitis."** **5. Revise #5 to read "acute or chronic bronchitis."** **6. Used in acute prostatitis.** **7. Pharyngitis should be added to criteria.**
Drug-condition Contraindications	Score: No contraindications OR drug justified despite contraindications: **Absolute:** Sulfonamides allergy, G6PD deficiency **Risk/benefit:** Blood dyscrasias or megaloblastic anemia, hepatic impairment, renal impairment	Flag justified = 5 Manual revision = 1 Practice factors = 1 Extenuating circumstances = 1 Criteria revision = 1 Total flags reviewed = 9 **1. Add to criterion #3, list 2: "unless there is evidence of reduction of usual adult dose."**
Drug-drug contraindications	Score: No contraindications OR drug justified despite contraindications: **Risk/benefit:** Oral antidiabetics, anticoagulants, phenytoin, methotrexate, thiazides	Flag justified = 5 Manual revision = 5 Criteria revision = 4 Total flags reviewed = 14 *1. Interaction between bactrim and oral hypoglycemic agents is rare.* **2. This was short-term sulfa use and documenting risk/benefit is unrealistic expectation.** **3-4. Potential interaction of thiazide and short-term bactrim not worth commenting on in records.**
Drug Monitoring	**If long-term, high-dose therapy:**	Flag justified = 3 Manual revision = 9 Criteria revision = 4 Total flags reviewed = 16

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Indicator	Criterion	Peer Review
	CBC q 1 mo	1. *Monthly CBC not justified for bactrim therapy.* **2. Monitor CBC q 6 mos.** **3. Document justification for CBC q mo. Patients are so variable with respect to general health status, no way to give rigid requirements for CBC frequency.** **4. CBC monitor q 6 mos for long-term use in sulfa.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (**bold**), and *Maryland* (*italics*).

TEST ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

BLOOD PRESSURE - DIASTOLIC BLOOD PRESSURE ≥90 **CODE: 01**

Indicator	Criterion	Peer Review
Test Indications		Flag justified = 1 Manual revision = 10 Practice factors = 3 Total flags reviewed = 14
Test Response	Overall score: Abnormal test results appropriately followed up	Flag justified = 8 Manual revision = 6 Practice factors = 1 Criteria revision = 4 Total flags reviewed = 19 1-2. No prior abnormal BP, thus, follow-up not required. **3. Should monitor whether patient has condition which causes a temporary increase in BP: pain, illness, or office hypertension.**
	Diast. BP 105-114 at least twice as new finding: Within 1 wk, treat as above OR drug therapy as below. Retest within 1 wk.	*4. For only 1 elevated diastolic >105, must review compliance, diet, weight changes in 2 wks.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

TEST ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

Hemoglobin/Hematocrit - Hgb >18 or Hgb <10 — Code: 02

Indicator	Criterion	Peer Review
Test Indications	Score: Test ordered appropriately for: **Conditions:** Anemia, blood loss, CAD, HTN, chron. renal failure, polycythemia, melena/stool-positive occult blood, fatigue/dizziness/conf., decreased tolerance for exertion or shortness of breath, infec. symptoms, aspirin use, NSAID use, gastrointestinal dis., concern re nutrition, self-care.	Criteria revision = 1 Total flags reviewed = 1 **1. CBCs justified to monitor drug known to affect hematopoiesis, i.e. chemotherapy.**
Test Response	Overall score: Abnormal test results appropriately followed up.	Flag justified = 7 Manual revision = 2 Extenuating circumstances = 2 Patient factors = 2 Practice factors = 1 Criteria revision = 3 Total flags reviewed = 14 1. Patient with nephrotic synd. and chronic renal failure. 2. Hgb 8.6 1/92 indicating anemia is chronic. **3. Need exception to doing a blood loss workup on patient taking chemotherapy. All test results should be responsibility of the ordering physician, in this case the oncologist.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

TEST ANALYSIS

Peer Reviewer Comments: Three State Combined Data

November 3, 1993

CHOLESTEROL-TOTAL CHOLESTEROL >240 **CODE: 03**

Indicator	Criterion	Peer Review
Test Indications	Score: Test ordered appropriately for: **Conditions:** Once q 2 yrs: Family hypercholesterolemia, IHD, HTN, diuretics (incr. LDL), B-blockers (decr. HDL), diabetes.	Manual revision = 3 Patient factors = 1 Decision rule = 1 Criteria revision = 9 Total flags reviewed = 14 *1. Any elderly person with any heart symptoms justifies a cholesterol.* *2. Should allow coronary insufficiency to = IHD* *3. Hyperthyroidism and hyperlipidemia need to be on list.* *4. Personal high cholesterol should be allowed criterion.* **5-8. Need to allow previous elevation or diagnosis of hypercholesterolemia.** **9. Routine SMA includes cholesterol.**
Test Response	Total chol. >240 and new abnorm: Within 3 mos retest, and if still ≥240, begin diet advice. Retest within 6 mos.	Flag justified = 7 Manual revision = 3 Practice factors = 1 Patient factors = 1 Criteria revision = 6 Total flags reviewed = 18 **1. Criterion needs to be expanded to allow for other appropriate actions taken to treat hypercholesterolemia. In this case, thyroid replacement was incr. and chol. was repeated in 4 mos.** 2. Diagnosis appears out of time span of study. *3. If triglycerides normal and HDL in 94-117 range and ratio total/HDL = 22-25, no need for intervention.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

TEST ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

POTASSIUM (SERUM LEVEL) K+ <3 OR K+ >5.7 CODE: 04

Indicator	Criterion	Peer Review
Test Indications		Flag justified = 1 Manual revision = 1 Total flags reviewed = 2
Test Response	Overall score: Abnormal test results appropriately followed up.	Flag justified = 10 Manual revision = 2 Criteria revision = 1 Total flags reviewed = 13 1.Case did not fit criteria. Absolute level would have been better utilized in test urgent response screen.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

TEST ANALYSIS

Test Data: Peer Review Comments

November 3, 1993

GLUCOSE-FASTING AND RANDOM BLOOD LEVELS-BLOOD GLUCOSE >140 IF FASTING AND NOT DM; >170 IF FASTING AND DM; >250 RANDOM

CODE: 05

Indicator	Criterion	Peer Review
Test Indications	Score: Test ordered appropriately for: **Annually if obese, family history of diabetes, or had gestational diabetes:** **Conditions:** Alcoholism, polyuria, polydipsia, polyphagia, wt loss, visual changes, glycosuria, DM, confirm self-monitoring values	Manual revision = 3 Practice factors = 1 Extenuating circumstances = 1 Criteria revision = 9 1. Ordering BS on prednisone therapy is appropriate. 2. BS part of chem. profile. 3. Criteria for justifying BS test not broad enough. *4. Indications should incl. hyperglycemia where sugar elevated on many occasions and patient on oral diabetic agent.* *5. Glucose justified as part of w/u for fatigue.* *6. Patient on micronase, glucose needs to be checked.* *7. Persistent diarrhea is another criteria indicating diabetes and reason to get bl. sugar.* **8. Hx of glucose intolerance** **9. Diabetic meds are justification for test.**
Test Response	Blood glucose >140 (if known to be fasting): Within 1 mo recheck before diagnosing DM, either with fasting glucose, glucose tol. test, OR 2-hr post-prandial test. Perform dipstick urinalysis for glycosuria/ketonuria. Review results within 2 wks.	Flag justified = 9 Patient factors = 1 Extenuating circumstances = 2 Manual revision = 1 Criteria revision = 3 Total flags reviewed = 16 **1. Allow glucose re-check in 3 mos for borderline glucose elevation of 140-200.**
	Bl. glucose >250 (not known to be fasting) or >170 (fasting): Within 1 wk start or incr. dose of oral antidiabetic OR start insulin therapy OR change insulin regime unless noted why action not performed (e.g., infec., stress, poor dietary compliance).	2. MD notes diabetes is followed by a primary care MD. **3. Add OR initiate dietary therapy.**

Inter-rater scores based on three-state data.

Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

TEST ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

CREATININE-SERUM CREATININE >1.5 CODE: 07

Indicator	Criterion	Peer Review
Test Indications	Score: Test ordered appropriately for: **Conditions:** Renal dis., HTN, diabetes, CHF, new onset proteinuria or abnormal urine cell count, ACE inh., nonsteroidal anti-inflamm. drugs, all drugs mainly excreted through kidneys.	Manual revision = 4 Patient factors = 1 Criteria revision = 4 Total flags reviewed = 9 *1. Test ordered as part of panel for other reasons.* *2. Expand list to incl. as 65 and older, use of aminoglycosides, renal artery stenosis, sepsis, crush injuries.* **3. Allow for labs done in panels.** **4. Expand criteria to include metastatic Ca and potential hypercalcemia.**
Test Response	New finding creat. 1.5-2.5: Within 1 mo, if on drugs that affect renal func., decr. or disc. meds. Check urine sediment exam and culture, and retest serum creat. within 3 mos. If still abnormal, request consult OR referral.	Flag justified = 7 Manual revision = 3 Extenuating circumstances = 1 Patient factors = 1 Criteria revision = 6 Total flags reviewed = 18 1. Creat. value of 1.6 in patient with no history of renal dis. on renal toxic drugs is not significant. 2. Creat. level 1.7 normal for this lab data. 3. Patient had creat 2.0 on 11/89 indicating PMH of chronic renal insufficiency. Criteria did not detect this as renal dis.; therefore, need better definition for renal dis. **4. Criterion should reflect normals from reference lab - in this case, 1.6 is within normal limits.** **5. Add phrase "and/or" after first sentence to allow changing drugs OR evaluation.**
	Creat. incr. >50% above that patient's prior baseline: Within 2 wks, if on drugs that affect renal func., decr. or disc. meds. Check urinary sediment exam and culture and creat. clearance. If creat. clearance has fallen, request consult OR referral within 2 mos.	6. Actions requested are too general. In most cases, only one action needed.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

TEST ANALYSIS

Peer Reviewer Comments: Three State Combined Data

November 3, 1993

ESR (ERYTHROCYTE SEDIMENTATION RATE) - ESR >730 CODE: 08

Indicator	Criterion	Peer Review
Test Response	ESR >30 and <80, new finding with no obvious cause: Within 1 mo, do routine history and physical exam only. Repeat test within 6 mos.	Manual revision = 1 Criteria revision = 2 Total flags reviewed = 3 1-2. Cause of ESR known.

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

TEST ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

Urine Dipstick for Glucose, Ketones, Protein, Blood, and Leukocyte Esterase (Marker of WBC)
≥3 + Glucose, ≥1 + Protein, 1 + Leukocyte Esterase, ≥1 + Heme Code: 09

Indicator	Criterion	Peer Review
Test Indications		Flag justified = 4 Manual revision = 4 Extenuating circumstances = 1 Total flags reviewed = 9
Test Response	≥3 + urine glucose in diab.: Within 1 wk check dietary and med. compl. and change therapy if required. If in non-diabetic, within 1 wk check bl. glucose at timed relationship to food intake, check urine ketones. Review results in 1 mo.	Flag justified = 8 Manual revision = 4 Extenuating circumstances = 1 Criteria revision = 6 Total flags reviewed = 19 1-2. Urine ketones not needed unless BS markedly elevated. **3. Follow-up after change in therapy should be lengthened to 3-4 wks.**
	≥1 + protein, new finding, no known diab.: Within 2 wks check urine cult. or leukocyte esterase. If infected, treat appropriately OR if not infected, within 2 wks check early a.m. urine sediment and dipstick and serum creat. If early a.m. proteinuria found, check 24-hr urine for protein. Review results in 1 mo.	4-5. 1+ protein in u/a from female w/no symptoms is not significant. Criteria need to begin at 2+ protein. **6. Re-check U/A in 1 mo.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

TEST ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

ECG

CODE: 10

Indicator	Criterion	Peer Review
Test Indications	Score: Test ordered appropriately for: **Conditions:** Not routine except with DM, HTN, CHF, coronary art. dis., intrinsic renal dis., valv. heart dis., arrhythmias, metabolic derangements, HTN, DM, PVD, preop eval., tricyclic antidepressants, CHF.	Flag justified = 2 Manual revision = 6 Criteria revision = 7 Total flags reviewed = 15 *1. Expand criteria to incl. synonyms (i.e., "chest pain," c/o angina).* *2. Expand criteria to incl. c/o palpitations or retrosternal chest pain.* *3. Incl. h/o chest pain over age 40.* **4-6. Chest pain should be indication for ECG.** **7. Syncopal episode needs to be added.**
Test Urgent Response	Score: Patients referred to hospital or ER for lab values: On ECG, new findings of acute ischemic changes, 2nd or 3rd deg. heart block, severe metabolic derangements, multifocal ventricular premature beats, or ventricular tachycardia.	Flag justified = 2 Manual revision = 2 Decision rule = 1 Criteria revision = 1 Total flags reviewed = 6 **1. If ECG changes are of uncertain duration and if patient asymptomatic, then immediate referral not needed.**
Test Response		Flag justified = 5 Manual revision = 9 Decision rule = 1 Total flags reviewed = 15

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

TEST ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

X-RAY: CHEST

CODE: 12

Indicator	Criterion	Peer Review
Test Indications	Score: Test ordered appropriately for: **Conditions: Annually in chronic smoker:** Known or suspected heart dis., chest dis., or esophageal dis.; heartburn/acid reflux; chest or shoulder pain; cough, fever, lung signs; cough and wt loss in smoker; preop chest x-ray; chest trauma, suspected rib fracture; known TB or positive PPD.	Flag justified = 4 Manual revision = 10 Criteria revision = 4 Total flags reviewed = 18 **1. Dyspnea or SOB are appropriate indications for a CXR.** **2-3. CXR justified for follow-up of certain cancers known to metastasize to lungs/chest.** **4. Add hypertension to criteria.**
Test Response	Pneumonia: Within 1 day treat with antibiotics and review within 1 wk. If symptoms persist, repeat x-ray and get PPD within 1 mo.	Flag justified = 2 Manual revision = 5 Criteria revision = 3 Total flags reviewed = 10 1. *PPD should not routinely be considered for all persisting pneumonias, but for those with clinically or radiographically suggestive TB.*
	CHF: Within 1 wk review salt intake. Start or adjust diur. therapy. Re-exam within 1 wk. If symptoms persist, augment treatment OR request consult OR referral.	2. *Dietary review of salt intake not paramount for an individual with chron CHF.*
	Lung infiltrate: Within 1 day treat w/antibiotics for >1wk, then repeat x-ray and get PPD after therapy if symptoms persist.	3. *Chron infiltrates should not be treated w/antibiotics until the exact nature is determined.*

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland (italics)*.

TEST ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

MAMMOGRAM CODE: 13

Indicator	Criterion	Peer Review
Test Response		Flag justified = 3 Extenuating circumstances = 1 Manual revision = 4 Total flags reviewed = 8
Test Indications	Score: Test ordered appropriately for: **Conditions: Annually:** Positive for family history of breast cancer, personal history of breast cancer, or screening.	Manual revision = 6 Criteria revision = 2 1. *Expand to incl. any palpable breast lesion.* 2. **Expand to include previous abnormal.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

PREVENTION ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

Indicator	Criterion	Peer Review
Blood Pressure		Flag justified = 8 Manual revision = 7 Practice factors = 1 Total flags reviewed = 16
Mammogram		Flag justified = 10 Extenuating circumstances = 1 Practice factors = 1 Total flags reviewed = 12
Breast Exam		Flag justified = 10 Extenuating circumstances = 1 Practice factors = 1 Patient factors = 2 Criteria revision = 1 Total flags reviewed = 15
Flu Shot Once in two-yr period	Score	Flag justified = 11 Manual revision = 2 Practice factors = 1 Extenuating circumstances = 1 Criteria revision = 2 Total flags reviewed = 17 **1. Patient refusals are not always marked in record.** **2. While the current recommendations should be promoted, they may be an insurmountable problem for the physician if patient does not get a flu shot.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, Iowa (bold), and *Maryland* (*italics*).

PROCEDURE ANALYSIS

Peer Reviewer Comments: Three State Combined Data

November 3, 1993

FLEXIBLE SIGMOIDOSCOPY CODE: 01

Indicator	Criterion	Peer Review
Procedure Indications	History of polyps; rectal bl./guaiac-pos. stool; signif. change in stool habits more than 1 mo; inflamm. vasculitic or granulomatous bowel dis.; unexplained diarrhea; prior colorectal cancer surg.; rectal exam suggests mass.	Flag justified = 3 Manual revision = 3 Criteria revision = 4 Total flags reviewed = 10 1. *Broaden criteria to include abdominal distension.* **2. Persisting more than 1 mo is too rigid.** **3-4. Screening every 3-5 yrs after age 50 needs to be added, per A.C.S.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

PROCEDURE ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

COLONOSCOPY CODE: 02

Indicator	Criterion	Peer Review
Procedure Indications	History familial polyposis synd., colon cancer; rectal bl./guaiac-pos. stools; signif. change stool habits more than 1 mo; inflamm., vasculitic, or granulomatous bowel dis.; unexplained diarrhea; prior colon cancer surg.; narrowing on barium enema; history benign polyps or sigmoid. finding neoplastic polyp.	Flag justified = 2 Manual revision = 2 Practice factors = 2 Total flags reviewed = 6 **1. Evaluation of chronic abdominal pain should be added.** **2. Add "evaluation of polyps seen on barium enema and palpable abnormalities by digital rectal exam."**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

FOLLOW-UP ANALYSIS
Peer Reviewer Comments: Three State Combined Data
November 3, 1993

Indicator	Criterion	Peer Review
Follow-up Hospital or consult	Overall score	Flag justified = 6 Manual revision = 4 Extenuating circumstances = 2 Patient factors = 2 Criteria revision = 4 1. MD note appears outside study time period. 2. Treatment of warts is an elective procedure.
	If consult requested, follow-up mention in record.	*3. Self-limiting problems should not be included in consult review.* *4. Passing remark regarding patient seen by cardiologist does not qualify as referral.*
	If hospitalization planned, follow-up mention in record.	
Follow-up of significant finding	Overall score	Flag justified = 3 Manual revision = 7 Decision rule = 1 Patient factors = 1 Criteria revision = 3
	If a mass is found, follow-up of finding in record.	**1. A simple problem such as sebaceous cyst is not a mass.**
	If a bleed is found, follow-up of finding in record.	
	If a weight loss is found, follow-up of finding in record.	2. Weight loss of 5 lbs not significant enough to initiate evaluation. **3. Need to quantify what is meant by significant weight loss. Five lbs is not significant.**

Inter-rater scores based on three-state data.
Peer Review data are combined Alabama, **Iowa (bold)**, and *Maryland* (*italics*).

Chapter 15
Evaluating DRRS and DCRS: Feedback of Results to Physician Groups

Abstract
As part of the evaluation of the claims profiling and record review, the preliminary results of the measurement were fed back to liaison physicians in each state. This process solicited feedback and comments about the perceived quality of the data and about potentially interesting findings. The meetings generated lists of possible CME topics and also of possible refinements to both the claims profiling and record review tools.

Meetings were held with physician groups in each of the three study states. The Maryland meetings occurred on October 21 and November 4, 1992. The Alabama meetings took place on January 28 and 29, 1993. The Iowa meetings occurred on February 25 and 26, 1993. Minutes of these meetings are provided as Appendices 1 (Maryland), 2 (Alabama) and 3 (Iowa).

Each meeting had four major objectives:

- Promote understanding of DEMPAQ methods;
- Present record review and claims profiling results;
- Communicate strengths and limitations of results; and
- Identify potentially interesting findings for further review, or possible CME.

The meetings in each state amply fulfilled our expectations of progress toward these objectives. The questions and comments of the physicians showed us that they clearly grasped the claims and records data as used in the DEMPAQ review method, despite the complexity of the results.

The format of each meeting was similar. The sessions opened with an overview of the DEMPAQ methods and philosophy (see attached appendices). Each meeting devoted a substantial block of time to an explanation of the DEMPAQ tools and feedback report formats. Examples of each type of report (claims and records) were explained in detail with time for questions. Next the results were presented, followed by a discussion of outstanding research issues.

Presentation of Results

Most of the time at each meeting was spent covering the results of the review. As the results were presented, the physicians were asked to judge whether the result seemed believable (based on the results of DEMPAQ reliability and validity studies) and whether the finding was potentially interesting from a CME standpoint. As described below, discussion was fruitful in these areas. In each state follow-up telephone calls were scheduled with physicians to discuss remaining results not covered during the meeting.

Strengths and Limitations of the Results

The discussions helped identify areas where the review instruments and review processes could be improved. The key changes and improvements suggested by the physicians are summarized in Appendix 4, Tables 1 (claims) and 2 (records).

Continuing Medical Education

The conversations with each of the physician groups began to highlight potentially interesting findings that might serve as CME topics. These included:

- Hypertension management and the use of newer antihypertensive drugs;
- Diabetes management and controversies about monitoring glucose control and monitoring for complications;

- Follow-up of abnormal test results;
- COPD diagnosis and use of bronchodilators;
- Ischemic heart disease diagnosis and admission for arrhythmias;
- Congestive heart failure;
- Prevention—especially breast exam and mammogram; and
- Antibiotic use.

Following the liaison briefing sessions, discussions about conducting CME workshops were initiated with appropriate individuals in each state. We met with Dr. James Hudson of the University of Maryland, Dr. George T. Smith from the University of Alabama School of Medicine and Mr. Lou Crist at the University of Iowa. Based on these discussions and the above list of possible CME topics, the following subjects were selected as the focus on the workshops in each state:

- Diabetes (all three states);
- Hypertension (all three states); and
- COPD (Maryland only).

Workshops were held during the summer and fall of 1993. A complete description of these workshops follows in the report on the Feedback and Education Modification to the Original DEMPAQ Contract.

Summary and Recommendations

Based on our meetings in each of the three states, we found that it was possible to work cooperatively with physicians for quality improvement using quality measurement of the kind developed by the DEMPAQ project. The reception of the data was thoughtful and sometimes challenging. Often the results provoked constructive and fruitful debate among the physicians as to the interpretation and importance of the findings.

The thoughtful reception of results was a reflection of the degree to which the physicians have been involved with the project. Over the last three years, physicians in each state have had the opportunity to: meet with the investigators and express concerns and reservations about the project and its methods; be involved with the development of the criteria and standards used in the review; and conduct peer review of the results. In addition, physician groups received regular project updates in the form of the *DEMPAQ Quarterly,* a quarterly newsletter prepared by project staff.

Should HCFA pursue further DEMPAQ type review, the proposed changes in Appendix 4, Tables 1 and 2 must be considered.

Appendix 1

MARYLAND: RECORD REVIEW MEETING
October 21, 1992

ATTENDING: Sheldon Goldgeier, MD, Christine Marino, MD, Abdul Nayeem, MD, Alfred Ossman, MD, R. Heather Palmer, MB, BCh, Jonathan Weiner, DPH, Howard Wilson, MD and Joseph Zebley III, MD; Jean Edwards, Sandy Moentmann, MA

The following describes the major topics of discussion at the October 21st meeting with the Maryland physician consultants. These notes are not intended to be exhaustive, but rather to highlight the main points of the evening.

General

In the early stages of the discussion, physicians were cautious about the results and often critical, but as they became more comfortable with the tone of the review, they began to participate more enthusiastically. The tone set was first of examining the credibility of the result in terms of the number of cases available, the inter-rater reliability for the item, and the results of peer review. As examples of extremes, less than ten cases available, inter-rater reliability falling below 80%, and a substantial majority peer review opinion criticizing the criteria, all obviously weakened the finding. As an example of peer review result weakening a finding, whenever a timing requirement was set at Q6-months, we had interpreted that fairly strictly in the Maryland abstraction process. Physicians were in agreement with the idea that Q6-month monitoring could be reviewed as twice within the year, at least 4 months apart.

As the meeting proceeded, physicians would frequently identify findings that seemed to indicate the opportunity for improvement, but almost always wished to conduct further review of the specific cases involved in order to clarify the issue and satisfy themselves as to the need for improvement.

Hypertension

Old diagnosis monitoring: the 6-monthly requirements were subject to the criticism of the decision rule adopted. The 1-year requirements -- physicians noted the high rate of failure to record fundoscopy and considered this an important finding. Also, the high rate of failure to perform urinalysis.

On *evidence for new diagnosis*, physicians agreed it would be important to adopt a new decision rule that a patient should be considered to have an old diagnosis of hypertension if the patient was on a drug that could lower blood pressure, provided patients did indeed have a new diagnosis of hypertension. The finding under the treatable causes indicator, that peer reviewers upheld at least half of the cases which failed to have investigation of underlying causes for a new diastolic blood pressure of 105 on two occasions, was thought to be an important finding.

Serum potassium as a test: Peer review generally supported the 9 failed cases, however detailed comments suggested again the algorithm was too complex and difficult to fit to real cases. Physicians seemed to feel the results important enough and the number of cases few enough that all should be reviewed in order to determine the type of quality problem and/or any need for criterion revision.

Ace inhibitors

The finding of failure to conduct *standing blood pressures* was expected but thought to be an important finding worth following. Monitoring of newly prescribed drugs: physicians agreed that the high not-met rates were of concern, although the criteria testing intervals were perhaps too strict. The requirement for a *blood pressure within 1 month* of the start of a newly prescribed ACE inhibitor was thought to be too strict, and that maybe 2 months should be allowable.

Alpha blockers

Again, *standing blood pressures* were an issue. Physicians were concerned at the 14% failure to record *dose and frequency*. The issue of 6-monthly or 12-monthly monitoring: the physicians felt that for 1-year monitoring requirements, the abstractor should be taught to look for monitoring plus or minus 2 months around 1 year, and that an acceptable alternative ought to be if the patient failed to keep an appointment.

Beta blockers

Drug/drug contraindications appeared to be a real finding: a 22% not-met rate highly upheld by the peer reviewers. The main problem identified by peer reviewers was the combination of calcium channel blockers with beta blockers. The *standing blood pressure* issue again emerged. Monitoring of newly prescribed drugs appeared also to be a problem. The physicians seem now to be critical of the notion of step therapy and to say that this was no longer the way hypertensive therapy was approached. There may need to be revision of this criterion. They felt that beta blockers were now considered step-one drugs, i.e, first line of treatment.

Centrally acting antiadrenergic drugs

Drug-drug contraindications Physicians felt that Timilol eye drops should be deleted as contraindications to the use of beta blockers. Some physicians disagreed with the contraindication given of calcium blockers, feeling that for some patients this was now a preferred combination.

Calcium channel blockers

Monitoring of newly prescribed drugs: apparently agreed to be a problem. Physicians felt this was an opportunity for education for quality improvement.

Loop diuretics

Drug monitoring: the 6-monthly requirements thought to be too tight.

Ischemic heart disease

Old diagnosis monitoring: weight gain appeared to be poorly noted; dyspnea was a problem with inter-rater reliability; there were no peer review results for this section.

Treatable causes: only 22 cases, with 15 who failed the criterion, not thought to be a big problem. The physicians felt that the criteria for finding treatable causes were not excessive, re asking that all of these be dealt with, however they felt that the issue of obesity could be handled by the recording of a weight. They were concerned as to know which of the elements were missing.

Serum cholesterol

The high failure rate for *followup of an abnormal cholesterol* of 240 or more. There was a suggestion for criterion revision. Overall though, there was concern that there was inadequate attention to abnormal cholesterol results.

Nitrate

Dose recording: relatively high not-met rate, but probably an issue of the various modes of administration of this drug. *Indications* potentially a problem. It was suggested that esophageal motility problems be added as an indication. *Monitoring of newly prescribed drugs*: evidence of education of the patient about angina had a high not-met rate, but was not particularly remarked upon by physicians.

Diabetes

The Maryland physicians believe that the failure to use *hemoglobin A1C* for monitoring of diabetics is a serious problem and should be addressed. They also felt the failure to record results of *fundoscopies* was a serious finding and should be addressed as a quality improvement problem.

Blood glucose

The *followup of abnormal serum glucose results*: physicians were concerned, even when there were occasional cases with delayed followup of very low or very high serum glucose, and suggested that those individual cases be specifically reviewed for continuing medical education purposes.

U/A dipstick

Followup of abnormal results: generally peer review upheld the failed cases, however time lines were thought to be too tight, for instance for following up proteinuria, comments seemed to reflect that a 3-month time interval was more sensible than 2 weeks and that a distinction should be made between male and female cases. There was a lot of discussion about the importance of following up leukocyte esterase cases and disagreement about the criterion. Again, for the finding of positive heme test, physicians thought it important to distinguish between the male and female cases. That seemed to be a general recommendation about the criteria for urinalysis that related to protein and blood in the urine.

Insulin

The issue of *blood glucose and hemoglobin A1C* again surfaced and is felt by the Maryland physicians to be of great interest. They feel that hemoglobin A1C is not being used as often as it should be, either for established users of insulin or newly prescribed insulin.

Oral antidiabetic agents

On the monitoring, again, the issue of *hemoglobin A1C and blood glucose* surfaced. And, it was noted from the detailed breakouts that use of hemoglobin A1C and performance of screening tests was surprisingly not better for insulin-dependent diabetics than for diabetics on oral treatment or no drug treatment. The only group who appeared to have better monitoring were those who had been on both insulin and oral antidiabetic agents during the study period, and the hypothesis was made that these were patients in whom there was a special difficulty over control and they were being switched from one drug to the other during the time period, and that may be why they were receiving better monitoring.

COPD

It was noted that *chest x-ray* may not be the right way to monitor, spirometry might be considered in a revision of criteria. *Evidence for new diagnosis*: peer review upholds the 80% failure rate on diagnosis of COPD.

Treatable causes: physicians felt it was a concern, that of 39 cases, 25 failed to have mention of smoking or an environmental precipitant of the symptoms.

CXR

The *followup of abnormal results* was felt to be a problem overall, despite the small numbers of cases in each type of chest x-ray abnormality. Physicians thought there was a pattern of poor followup of abnormal results of chest x-ray.

Bronchodilators

Dose and frequency: there were rather high rates of non-documentation, but this was felt to be an issue for manual revision, since the drugs are given by inhalation. Drug indications: the physicians were concerned by the 26 of 44 cases who appeared not to have the drug appropriately prescribed and felt that this issue should be reviewed. Monitoring: the high failure rates for *monitoring of an established drug* and of newly prescribed drugs, it was thought the criteria were probably too strict, that a longer time interval of 1 month plus or minus 2 weeks or even 8 weeks, and that phone contact rather than an actual examination were probably adequate.

The physicians had more concerns about oral bronchodilators than inhaled bronchodilators.

Cholinergic inhibitors

Indications: physicians apparently felt that they should now be considered a first-line drug, not used only if beta antiadrenergics and xanthines fail.

Xanthine derivatives

Dose and frequency: the failure to record dose and frequency in about 6% of cases was thought to be a very important finding. *Indications*: the 15% of cases without appropriate indications was felt to be a serious issue. In general, the physicians felt that theophylline is a problem drug and it should be much less used. *Drug indication/contraindications*: physicians now felt that congestive heart failure should be omitted from the list of risk/benefit conditions, that is that it is acceptable to use xanthines in patients with congestive heart failure. *Drug/drug contraindications*: there was a great deal of debate as to whether furosemide and beta blockers were truly contraindications, but the 31 of 116 cases who apparently had drug contraindications we thought worthy of further investigation. *Monitoring of pre-existing drugs*: because low dosage theophylline is used, physicians thought that only annual monitoring would be adequate. *Monitoring for newly prescribed theophylline*: the 1-week limit was thought too tight, and the 2-3 week timing acceptable. In general however, theophylline was considered a drug whose use should appropriately be waning and therefore this issue is worthy of review.

Prevention

The key issue with these results is of course that Medicare has not paid in the past for mammograms and flu shots and the high not-met rates for these two items are probably not real findings. The high not-met rate for breast exam is much more likely to be a real finding and was upheld by peer reviewers.

Appendix 2

ALABAMA: CLAIMS PROFILING MEETING
January 28, 1993

Attending: Robert Sherrill, MD, Nadine Schiesz, RN, Laurie Hall, MD, George C. Smith, Jr., MD, Pink Folmar, MD, Brian Perry, MD, Peter Morris, MD, Deborah Garnick, ScD, Jean Edwards, R Heather Palmer, MB BCh, Ann Lawthers, ScD

Dr. Palmer opened the meeting with a brief overview of the history and goals of the DEMPAQ project. Dr. Garnick then proceeded to go through the results of the various profiling efforts and to elicit reactions to the material.

The major comments of the evening are summarized in the paragraphs below. The material has been organized by profile, with general, overall comments first.

General comments:

Several physicians suggested that a single snapshot claims profile was hard to interpret, in part because it is possible to hypothesize alternative explanations for most individual findings. But with repeated measures, the data may become more interesting.

At several points in the evening, it was commented that an urban/rural analysis would be extremely useful and relevant. In the minds of several physicians, an urban/rural breakdown would probably be more helpful than a specialty division, although a specialty breakdown was of some interest.

Diabetes

The following findings elicited comment and alternative explanations: HgbA1c, urinalysis, optometry, blood sugar, and admissions.

It appeared to be the general consensus that while it is possible to agree with the *Hgba1c* standard in theory, in practice one discovers that the test is not needed. Patients who are well controlled on diet do not need HgbA1c. (Note discrepancy between the views of the Alabama physicians and the views of physicians in the other two states.)

Several items included in the profile may not be billed to Medicare. A low rate of *urinalysis* may be because the physician does the test in his office and then does not bill separately for the item. It is included in the general office visit charge. As a result of CLIA, physicians may now perform tests in the office but not bill for them. *Optometry* visits that take place in an optical shop may not be billed to Medicare. Likewise, *blood sugar* monitoring may not appear in the claims because of the use of Acucheck at home.

The rate of *admissions* for hyperosmolar coma seemed low. It was suggested that perhaps the diagnosis is coded as sepsis or something else.

Hypertension

For hypertension the physicians felt that the standard of appropriate care may vary by specialty. Therefore, they were interested in seeing specialty comparisons.

The use of *lipid screening* was debated. It was suggested that perhaps these tests shouldn't be done in the elderly (or at least the old elderly) because it takes so long to effect a clinically important change in lipid levels. One suggestion was to stratify these items by age.

It was suggested that perhaps *creatinine* should be moved to recommended care.

The *hospitalization rates* for CVA, hyperkalemia and hypokalemia appear low. The physicians would have expected more hypertensive patients to be hospitalized with these conditions. Perhaps these complications are coded as something else, such as arrhythmias, renal complications or drug complications.

COPD

Several doctors noted that it is difficult to get Medicare to pay for *home O2*, and therefore it is not surprising to see so many people without home O2.

Patients who see a pulmonary specialist are more likely to get *spirometry* than patients who see primary care physicians.

The number of patients with 5 or more *CXRs* seemed reasonable for a population of COPD patients who are prone to lung infections.

CHF

The finding of 1.8% of patients with five or more *echocardiography* exams is surprising. The consultants wonder if perhaps tests are being repeated unnecessarily, or if there is a data error.

The consultants were surprised that only 2.6% of CHF *visits were to cardiologists* and only 10% of patients saw a cardiologist. They advanced the theory that a patient is admitted to the hospital with CHF, treated in hospital by a cardiologist and then discharged to the primary care physician. These in-hospital visits would not show up.

IHD

The issue of age breakdown was revisited. It was suggested that since most geriatric studies use 65 - 79 and 80+, we might want to do the same.

The finding of some patients with five or more *cardiac catheterizations* in a year seems like an odd finding. Why would any patient have more than 5 cardiac catheterizations?

Office practice profile

It was suggested that *table 6, "Distribution of your primary patients by type of co-morbidities*" is "at the mercy of how many diagnoses are coded." If the physician fails to code completely then the ACG assignment may be off. This in turn could lead to misleading interpretations of the data.

Reliability and validity

There was some skepticism about the reliability of the diagnosis data, particularly if physicians only coded diagnoses that justified the ancillary services. Thus the project would miss some patients with our rules for including patients in a profile.

At one point in the evening, one of the physicians noted that prior to RBRVS, many GP's tended to code all their visits in one way, for example "brief," regardless of how long or extensive the visit really was. This would bias the office practice profile results because the RU's would understate the true RUs for the patient.

ALABAMA: RECORD REVIEW MEETING
January 29, 1993

ATTENDING: Robert Sherrill, MD, Nadine Schiesz, RN, Carlene Gibbons, Barbara Baites, Lauri Hall, MD, Pink Folmar, MD, Peter Morris, MD, Brian Perry, MD, George C. Smith, Sr., MD, George T. Smith, MD, Steve Boger, MD, R Heather Palmer, MB, BCh., Ann Lawthers, Sc.D., Nick Wareham, MD,

Dr. Palmer opened the meeting by explaining the formats used to display the records data and then proceeded to go over each item reviewed by the project, soliciting feedback.

The major comments of the evening are summarized in the paragraphs below. The material has been organized by broad category, overall comments first.

General

It was pointed out that the notation in the chart vis a vis awareness of an absolute or risk/benefit contraindication would usually occur the first time the drug was prescribed. Subsequently, the physician may speak with the patient about the contraindication, but rarely would s/he make a new notation in the chart.

Antibiotics

It was the general consensus that the *indications* list for antibiotics was too restrictive. The standard of care for use of the various antibiotics in practice is looser than suggested by the criteria. In order to develop appropriate standards for antibiotics one needs knowledge of community acquired infections and common infections in the community.

Recording the *dose and frequency* of antibiotics seem to be a potential problem (50% to 70% of the time it was recorded, as compared to 90% or better for anti-hypertensive medications). This might be target for education.

Corticosteroids

It was suggested that the criteria for *drug-drug contraindications* required revision. Corticosteroids and NSAIDs should listed as a risk/benefit contraindication rather than an absolute contraindication.

Requiring a *PPD* prior to starting a high-dose, long-term corticosteroid may be an appropriate educational target. The DEMPAQ data on this issue are somewhat soft but the sense was that many physicians do not get a PPD prior to starting long-term corticosteroids.

Hypertension

The consultants commented that physicians usually do not record *diet monitoring*. They tend to be more concerned with weight. The question then becomes, "Is it important to monitor diet? Should it be a one time instruction?" One individual commented that drugs do such a good job of controlling blood pressure that a low salt diet no longer makes a difference.

With respect to *fundoscopy*, several physicians commented that if the blood pressure is well controlled, they will not do a fundoscopy. It seemed to be the consensus that fundoscopy was more important for monitoring malignant hypertension and that this condition was not particularly prevalent in the elderly. Fundoscopy was also deemed a vital part of a new diagnosis workup. The group came close to suggesting that two sets of criteria were needed: new diagnosis workup and old diagnosis monitoring.

Serum potassium

It was noted that "red herring" results sometimes come from test panels. A physician may have ordered a panel for a test other than potassium, the potassium came back abnormal and was not acted upon. It was suggested that in ambulatory care it is easy to miss abnormal results. This is an issue where re-organizing one's practice may enable abnormal results to be identified.

Ace inhibitors

Standing blood pressure checks clearly are not done by primary care physicians. Several physicians commented on the hassle factor of doing standing BP checks. They noted that when ACE inhibitors first came out, standing pressures were recommended, but that recommendation no longer appears. The consensus seemed to be that standing BP checks are indicated if the patient has symptoms suggestive of postural hypotension.

Diabetes

Two changes to the criteria were suggested. First, it was noted that a *comment on glucose control* was unlikely to be included in the record and furthermore that failure to comment on glucose control was not a significant finding. The second change involved the *indications criteria* for a new diagnosis of diabetes. It was suggested that the trigger value for a random glucose should be higher than 250.

Insulin

Several of the consultants felt that there was no need to do a *HgbA1c* if the patient appeared to be well controlled, if the patient was a type II diabetic, or if the Hgba1c results were not expected to change therapy. It was further noted that HgbA1c is an expensive test. There was acknowledgement that younger physicians are now taught to use it.

Oral hypoglycemics

With respect to *drug-drug contraindications*, the criteria should delete the following two drugs from the list of contraindications: insulin and beta blockers. Both these agents can be used with oral hypoglycemics.

COPD

The determination of a new diagnosis seemed to be a potential CME target. That is, the consensus was that perhaps COPD is over-diagnosed.

Prevention

Mammogram and breast exam both had a low percentage of cases where the criteria were met. It was the impression of the consultants that this finding was probably real. The low percentage of patients receiving a flu shot, on the other hand, is probably an artifact. Notation of a flu shot is often missing either because the patient does not come into the office in the fall or because the patient receives a flu shot elsewhere.

Appendix 3

IOWA: CLAIMS PROFILING MEETING
February 25, 1993

ATTENDING: James Maroc, MD, IFMC, John Carroll, MD, Kevin Quinn, MD, Milt Vangundy, MD, Larry Severeidt, MD, Dana Shaffer, DO, Kevin deRegnier DO, Robert Good DO, Nancy Paine, IFMC, Becky Hemann IFMC, Sherri Segebart IFMC, Donna Bottroff IMS, Jay Rosenburger DO, Jenny VanGundy, IFMC, Jinnet Fowles, Heather Palmer, Ann Lawthers

The following describes the major topics of discussion at the Thursday, February 25, 1993 meeting with Iowa physicians. These notes are intended to capture key points and therefore are not exhaustive. The comments are organized by clinical item.

General

The issue of categorizing primary care physicians received some discussion. It was the general consensus of the group that there should be three groups: General Practice, Family Practice, and Internal Medicine. Osteopathic physicians would not be broken out as a separate group. Rather than using the provider's self designated specialty from the provider file, it was suggested that Board Certification should be used to identify a physician's specialty.

The usefulness of the claims profiles surfaced as a topic of concern. The question was raised whether it is possible to tell from the claims whether a physician is taking satisfactory care of the patient. The claims reveal whether a test is done, not whether the physician followed up the result correctly. Another issue was how to determine a clinically meaningful number. The differences between the two groups may be statistically significant but not clinically important.

Diabetes

Among recommended care items, HgbA1C, triglycerides and HDL generated comment. For table 2, admissions for complications, several revisions were suggested.

HgbA1C provoked considerable discussion. The tone and tenor of the discussion was thoughtful and ranged from whether annual or six monthly HgbA1C was a valid standard, to the notion that not enough testing is being done, to the idea that the clinician may not be able to really intervene in a patient who is not well controlled. The discussion concluded that it may be useful to identify CME topics based on areas where practice is changing.

Triglycerides and HDL were viewed as not particularly useful recommended care items. It was pointed out that these tests may be done less frequently if the patient has had previously normal results.

It was suggested that *AMI and stroke* may be more useful measures of admissions for complications than the list currently included in the profile. An alternative explanation for the low rate of hypoglycemic admissions may be because such patients are held overnight for observation and never admitted.

In measuring the *length of time from discharge to ambulatory visit* for CVA, it was suggested that we need to look at nursing home visits.

Hypertension

There was considerable discussion around the clinical usefulness of the various profile components. One physician suggested moving cholesterol to recommended and potassium to limited use. Another suggested the importance of U/As to monitor proteinuria and creatinine for renal function. Overall, the sense was that only office visits were truly of interest for hypertension. Even trended data was not particularly useful because there were so many other factors, such as drug therapy, that are more important. Thus the hypertension profile appeared to yield little in the way of CME targets.

The post discharge information appeared to be valid. Most of it fell into the "see me in three or four weeks," range. Several participants suggested that trending hospital data may be interesting and helpful if it reveals a target that is amenable to education.

Congestive heart failure

Several alterations to the profile components were suggested. First, is was suggested that *flu shot* be eliminated because they rarely appear in the claims file. Next, a stratification for the limited use items of *contrast angiography* and *exercise stress test* was suggested. Patients could be divided into those with ischemic heart disease and those without.

Dig toxicity was proposed as an admission diagnosis that would be very interesting to track under *"Admissions for Complications."*

Ischemic heart disease

Several consultants suggested that the limited use items be stratified by age because resource use typically varies depending on patient age. For example, an 85 year old patient with IHD may not be subjected to a treadmill test.

COPD

Because the use of home O2 is so tightly regulated, it was viewed as a good stratifier.

Osteoarthritis

The admissions for OA of 5.74 was a surprise. The admission rate for bleeding ulcer appeared consistent with the literature and a good indication of a serious complication of therapy.

Office practice profile

The information here looked very interesting, however its use and importance was the subject of some debate. "What do you do with this? Is this good or bad? Where do you go from here?" were among the questions asked.

IOWA: RECORD REVIEW MEETING
February 26, 1993

ATTENDING: James Maroc, MD, IFMC, John Carroll, MD, Kevin Quinn, MD, Milt VanGundy, MD, Larry Severeidt, MD, Dana Shaffer, DO, Kevin deRegnier DO, Robert Good DO, Nancy Paine, IFMC, Becky Hemann IFMC, Sherri Segebart IFMC, Donna Bottroff IMS, Steven Craig, MD, Jenny VanGundy, IFMC, Jinnet Fowles, Heather Palmer, Ann Lawthers

The following describes the major topics of discussion at the Friday, February 26, 1993 meeting with Iowa physicians. These notes are intended to capture key points and therefore are not exhaustive. The comments are organized by clinical item.

General

Among the themes raised during the discussion were two general ones: the likely use of the results, the usefulness of the results.

Three possible uses of the data were discussed: (1) to reassure those who pay so that they know they are getting what they paid for, i.e. accountability to the payor and the public; (2) detect overall patterns that are a problem, i.e. put into the hands of the medical profession the tools to conduct self monitoring and review; and (3) to detect individual instances of problematic care. The DEMPAQ team reiterated that the primary focus was on the usefulness of the results to the medical profession.

Related to the use of the data was the theme of usefulness. Throughout the discussion, the consultants continually brought the conversation back to whether a particular finding was important from a CME perspective. This development was quite encouraging as it indicated that the intent and thrust of the review had been grasped.

Hypertension

The items that received the most discussion were: monitoring diet, fundoscopy, U/A dipstick for protein and "Evidence for a New Diagnosis."

Diet as a monitoring item was viewed as not particularly important. Most felt that doctors were likely to discuss diet with patients but not document the discussion.

The issue of *fundoscopy* generated debate. Some suggested that primary care physicians may have difficulty doing such exams and therefore ophthalmologists were best qualified to do the exam. Others commented on its importance for newly diagnosed hypertensive patients. The discussion was summed up by the statement "We apparently aren't doing fundoscopic exams.

Is this a good CME target?"

The general consensus on the next item, the *U/A dipstick for protein*, was that this finding (37% "fail") was potentially important. Physicians should monitor for protein.

Two comments were made in relation to *"Evidence for a New Diagnosis"* First, it was suggested that the SBP level indicating hypertension should be increased to > 160. The second comment was that the review system needed a way to track down missed diagnoses. Dr. Palmer pointed out that the Test Function was designed, in part, to serve this purpose.

Ace inhibitors

Standing blood pressures were the primary topic of conversation for ACE Inhibitors. In the absence of symptoms, there was some question about the importance of doing a standing blood pressure. This generated quite a discussion about how to organize one's practice so that standing BP's get done. Each of the physicians present had a slightly different routine for making sure that a patient had the necessary monitoring.

Ischemic heart disease

It was suggested that the *evidence for a new diagnosis* needed to be modified somewhat. Epigastric pain based on exercise might be accepted. Or perhaps referred pain. It was the general feeling that this was not a CME issue but rather an issue for criteria revisions.

Diabetes

Pulse and fundoscopy were the two major topics of discussion under diabetes. It was felt that pulse should be identified as a separate issue. *Pulse* was reaffirmed as an important monitoring action because diabetic patients are prone to heart disease and other complications. As in hypertension, *fundoscopy* generated comment. Again the issue was whether primary care physicians should be monitoring the fundus. However, the researchers pointed out that referral for fundoscopy also met the criterion.

It was pointed out that the percent acceptable performance diabetes monitoring was between 20 and 55 percent while for ischemic heart disease, the performance was in the range of 50 to 90%. The question was raised, "Are diabetics being treated less carefully than IHD patients? If so, does this make clinical sense."

U/A dipstick

Some discussion centered on whether the trigger value for review should be changed from *1+ protein* to 2+ protein. The consensus appeared to be that a 1+ protein in diabetics was important, while in a non-diabetic, a 2+ protein was perhaps the more appropriate cut point. It was noted that *leukocyte esterase* is not on all brands of dipstick.

COPD

Much discussion focused on the *criteria for making a diagnosis*. It was the general consensus that physicians occasionally label a patient has having COPD without doing pulmonary function tests first or meeting the diagnostic criteria based on history and physical. Such a doctor may institute a trial of medication for COPD to see how the patient responds. It was felt that rural physicians in particular would be less likely to have spirometry equipment in the office and therefore less likely to make a COPD diagnosis based on PFTs. It was suggested that perhaps this subject, the diagnosis of COPD would make a good topic for "informed discussions."

Antibiotics

Several comments surfaced in relation to the indications for using antibiotics.

The consultants were quite explicit that is inappropriate to use antibiotics for upper respiratory infections. It was suggested that perhaps physicians who say they prescribe antibiotics for URI may use the term to refer to pharyngitis and sinusitis. For patients with COPD on the other hand, the consensus was that it was appropriate to prescribe antibiotics at the beginning of a viral infection.

Page 85 of the criterion book generated the comment that an immediate allergy to penicillin should only refer to an anaphylactic reaction, not a rash.

The indications for cephalosporins need to be broadened. They are indicated for a variety of first line treatments, not just for ailments where a penicillin allergy already exists or the organisms are proven insensitive to other antibiotics.

Antibiotics also generated an important discussion about the sources used to document the criteria. The DEMPAQ project used Drug Evaluations as the primary reference for drugs. The Iowa physicians pointed out that this was not a readily available resource in most offices. The PDR was much more common. The question arose, if one does not have access to the Drug Evaluations then should one be held to a standard based on the book?

Appendix 4

Chapter 15: Evaluating DRRS and DCRS: Feedback of results to physicians groups

Table 1

Liaison Physicians Suggestions for Revising DEMPAQ Claims Tool

Profile/Component	Suggestion
Lab tests	As a result of CLIA, physicians may perform a variety of tests in the office but not bill for them.
Diabetes: ophthalmology visits	Missing optometry visits that take place in an optical shop and not billed to Medicare.
Diabetes: hyperosmolar coma	Perhaps this is coded as sepsis. Should check the coding.
HTN: lipid screening	Stratify by age since these tests shouldn't be done for the old elderly. It takes a long time to effect a clinically important change in blood levels.
HTN: creatinine	Move to recommended care.
HTN: hyper-hypokalemia	Perhaps coded as arrhythmias, renal complications, drug complications
Diabetes: HDL and triglycerides	Delete. These are not particularly useful recommended care items.
Diabetes: hospitalization	AMI and stroke are more useful measures of admission than measures currently included in profile.
Hospitalization	Present trends over time. It may be easier to pick up targets for education by looking at trended data.
CHF: angiography	Stratify by IHD, those with and those without. Current stratification, all CHF and the subset of those with IHD is *not* helpful.
CHF: stress test	Stratify by IHD, those with and those without. Current stratification, all CHF and the subset of those with IHD is *not* helpful.
CHF: hospitalization	Add dig toxicity.
IHD: process measures	Stratify by age. The over 80 group is not likely to receive treadmill tests, etc.
COPD: process measures	Stratify by home O2. Home O2 is tightly regulated.

Table 2

Liaison Physicians Suggestions for Revising DEMPAQ Records Tool

Clinical Item/Indicator	Suggestion
Dx and Rx:	Add an acceptable alternative of "patient failed to keep appointment."
Dx and Rx: monitoring	Change the "q. 6 months" monitoring to • twice a year, at least 4 months apart; or • within 6 months (plus 1 month, so in effect within 7 months); or • every 6 months, plus or minus 2 months.
Dx and Rx: monitoring	Change "q. 12 months to • within 1 year; or • every 12 months, plus or minus 2 months.
Dx: COPD, evidence for new Dx	Should the criteria accommodate physicians who may do a trial of medication to see how patient responds and call this COPD? What about rural physicians who are less likely to have spirometry equipment in the office? Or is this a problem of distinguishing between "old" versus "new" diagnosis with the review criteria?
Dx: COPD, monitoring	Delete CXR requirement Add spirometry requirement
Dx: Diabetes, new diagnosis	Trigger value should be a random of greater than 250.
Dx: HTN, fundoscopy	Only important if patient has malignant hypertension, or if patient has new diagnosis.
Dx: HTN, diet monitoring	Diet monitoring is not that important. Low salt diet is no longer an issue with the advent of new drugs.
Dx: HTN, evidence for new dx	The SBP level indicating HTN should be increased to > 160.
Dx: HTN, new diagnosis	Perhaps should add an indicator, "appropriate new diagnosis workup" which would include fundoscopy for HTN.

Dx: HTN, old versus new	Allow an "old" diagnosis finding if the patient is on anti-hypertensive medication prior to the mention of HTN. Also allow an "old" diagnosis finding for patients on anti-hypertensive drugs, even if the physician neglects to mention *hypertension* anywhere in the record.
Dx: IHD, evidence for new dx	Accept "epigastric pain on exercise," or "referred pain."
Dx: Treatable causes	Asking for the reviewer to find all of the listed actions is an excessive requirement. The review should permit each cause/co-morbidity to be scored separately. This would allow identification of which actions are not done by primary care physicians.
Prevention: Flu shot	Need acceptable alternative, "Patient did not come in for visit in fall."
Rx: Antihypertensive, monitoring	Standing BPs are important, *only* if patient complains of orthostatic hypotension.
Rx: Bronchodilators, monitoring	Time frames too strict. Should be 2 months. Phone contact should be adequate monitoring.
Rx: Bronchodilators	Revise the manual to specifically accept "two puff" as dose.
Rx: Corticosteroids, contraindications	NSAIDs are a risk/benefit contraindication rather than an absolute contraindication.
Rx: ACE Inhibitors, drug monitoring	Time frames for new drug monitoring are too strict, particularly, a BP within 1 month. A BP within 2 or 3 months is more reasonable.
Rx: Anti-hypertensive drugs	Stepped therapy is no longer "in vogue."
Rx: Antibiotics	Indications lists are in general *way* too restrictive. However, there was disagreement among physicians in the various states as to the appropriate use of antibiotics. Physicians in one state called for the use of antibiotics with upper respiratory infections and physicians in another state saying it is clearly inappropriate to use antibiotics in these circumstances.
Rx: Antibiotics	Make it clear that it is appropriate to prescribe antibiotics for COPD patients at the beginning of a viral infection.
Rx: Beta Blockers	Beta blockers are appropriate first line drugs for treatment of HTN.

Rx: Beta Blockers	Calcium channel blockers should be deleted as contraindication to beta blockers.
Rx: Beta Blockers	Timilol eye drops should be deleted as contraindications to the use of beta blockers.
Rx: Cephalosporin	Indication list needs to be broadened. They are indicated for a variety of first line treatments, not just ailments where penicillin allergy already exists, or where the organism are proven insensitive to other antibiotics.
Rx: Cholinergic inhibitors, indications	These should now be considered first line drugs, not used only if beta adrenergic and xanthines fail.
Rx: Monitoring new drugs	Evidence of education not really that important. Perhaps this criterion should be deleted.
Rx: Nitrate	Add esophageal motility problems to list of indications
Rx: Oral hypoglycemics, contraindications	Eliminate insulin as a contraindiation. This agent can be used with oral hypoglycemics.
Rx: Oral hypoglycemics, contraindications	Eliminate beta blockers as a contraindiation. This agent can be used with oral hypoglycemics.
Rx: Xanthine, contraindications	Remove heart failure from list; it is no OK to use with CHF patients.
Rx: Xanthine, contraindications	Remove beta blockers from list of contraindications.
Rx: Xanthine, contraindications	Remove furosemide from list of contraindications.
Rx: Xanthine, monitoring	1 week time limit is too tight.
Test: Cholesterol	Revise criteria for follow-up of an abnormal cholesterol of 240 or more.
Test: Response	How important is it to follow-up abnormal results that are obtained incidentally, e.g. as part of a panel? Should the criteria include an acceptable alternative - "test results incidental, part of panel," to excuse not following up certain abnormal results?

Test: Test response	The follow-up sequences are too complex. Either • simplify, or • "disentangle" the sequences, that is, score each follow-up action separately.
Test: U/A	Response time frames too tight.y
Test: U/A	Develop separate criteria for male and female cases.
Test: U/A, trigger value	The trigger value for review of *non*-diabetics should be changed from 1+ protein to 2+ protein.

Chapter 19
Variation in Office-Based Quality: A Three-State Comparison of Care Provided to Medicare Patients with Diabetes and Congestive Heart Failure

Abstract
This chapter presents a practice variation analysis relying on the database we developed for DEMPAQ's claims profiles. One objective of this chapter is to offer a cross-state comparison of the degree to which quality-related parameters were met for two of the most common conditions—diabetes and congestive heart failure (CHF). A second objective is to identify and discuss physician, patient, and geographic factors associated with higher or lower compliance with recommended standards of care.

The main source of data for this analysis was Medicare claims submitted during the period July 1990 to June 1991 on behalf of the aged Medicare beneficiaries residing in the three study states. From these data, we identified patients who had made one or more office-based medical visits where the physician had noted a primary diagnosis on the claim for either diabetes or CHF. Across the three study states, approximately 10.9% of the DEMPAQ study population were thus identified as diabetics (n=97,388) and 8.0% were found to have congestive heart failure (n=71,496).

The dependent measure that was the focus of analysis was the proportion of in-scope patients within a primary care practice who received one of eight procedures indicative of quality (e.g., HgbA1C for diabetics) at least once during the 12 month study period. The independent variables included the following practice attributes: type of practice (e.g., specialty, group or solo); location of practice; patient age and sex; patient commodities; and practice size and case load.

We used linear probability regression (OLS) models to estimate the relationships between practice characteristics and procedure use. We also applied arc sine transformations and TOBIT models.

The findings of this study indicate that clinical actions in the primary care sector exhibit fairly wide levels of practice pattern variation. This variation appears to be not only closely associated with patient related factors, such as case-mix, but also to characteristics that are linked directly to the physician (e.g., specialty) and geography (e.g., location of practice).

For those clinical activities that are accepted as "ideal standards" of care (e.g., HgbA1C or eye exams among diabetics) this study provides some evidence that quality improvement programs should be targeted more directly at certain types of physicians (e.g., general practitioners) in particular geographic areas (e.g., rural locales, or selected states). For other procedures where general wide-scale use is not deemed effective (e.g., blood glucose for diabetics or chest x-rays among CHF patients) this study indicates that certain provider and geographic factors are associated with potential overuse of the services. This aspect of the study underscores the need for better scientifically derived clinical practice guidelines and programs to implement them.

In addition to its methods and findings, this chapter will have implications for those charged with improving the quality of care within the Medicare program. Even though similar primary care practice oriented analyses have been recommended for years, to the best of our knowledge, this is the first time that such analyses using Medicare claims data has been completed.

Introduction

The overarching goal of the DEMPAQ project was to develop and test methods that could be used to improve the quality of primary care delivered to Medicare beneficiaries. As part of this project, HCFA's insurance claims data were used to develop comprehensive profiles of the current patterns of practice of primary care physicians in three states (Maryland, Iowa, and Alabama). One key intent of these profiles was to document the degree to which standards of quality were attained in the delivery of office-based care provided to Medicare beneficiaries under treatment for common chronic conditions. This information (along with data derived from the medical record review portion of the DEMPAQ study) was meant to be used by local and regional physician groups as one important input into the quality improvement process.

In Chapters 6, 7, and 8 of this report, copies of condition-specific claims-based profiles are presented by state, specialty, and disease category; this chapter will not reiterate these results. Rather, given the general lack of knowledge regarding office-based patterns of quality, the goal of this chapter is to provide a summary and synthesis of these findings.

This chapter's first objective is to offer a cross-state comparison of the degree to which quality-related parameters were met for two of the most common conditions: diabetes and congestive heart failure. Its second objective is to identify and discuss physician, patient, and geographic factors associated with higher or lower compliance with recommended standards of care. This chapter will have implications for those charged with improving the quality of care within the Medicare program.

Even though similar primary care practice oriented analyses have been recommended for years, this is the first time that such an analysis using Medicare claims data has ever been attempted. Therefore, it is hoped that the methods and findings reported here will be relevant to those interested in understanding factors associated with variation in patterns of primary care quality.

Methods

Sources of Data

The main sources of data for this analysis were the Medicare claims submitted during the period July 1990 through June 1991 on behalf of the aged Medicare beneficiaries residing in the three study states. HCFA's HISKEW beneficiary eligibility file was used to identify an at-risk denominator for the study. The data were derived from HCFA's new 100% National Claims History File (NCHF). The methods used to extract and prepare these data are described in Chapter 5 of this report.

From the NCHF, we identified patients who had made one or more office-based medical visits where the physician had noted a primary diagnosis on the claim for either diabetes or congestive heart failure (CHF). Across the three study states, approximately 10.9% of the DEMPAQ study population was thus identified as diabetics (n = 97,388) and 8.0% were found to have congestive heart failure (n= 71,496).

To perform the analysis, a specialized database was constructed from several sources. After the target populations were identified in each state, 100% of claims submitted for services delivered to that person during the 12-month period were selected and summarized. The next step of the database development involved the aggregation of patient-specific records to the level of a medical practice, which was the main unit of analysis for this practice variation study.

We developed an algorithm to identify a beneficiary's primary care source (PCS). We defined a PCS as the primary care physician (or group) specializing in general practice, family practice, or general internal medicine, who has seen a particular beneficiary more frequently than any other primary care provider. In developing this PCS assignment we considered only face-to-face encounters in an office or clinic setting. Ties (which occurred less than 6% of the time) were broken by assigning the patient to the physician practice providing more intensive services as characterized by the relative value of all visits and procedures billed by that practice.

To be included in this analysis a primary care source had to have an overall total of at least 25 Medicare patients so assigned. To ensure that the physician was continuously practicing medicine during the year, we further restricted the PCS designation to providers practicing within the state submitting claims during each of the year's four quarters.

Over 75% of each state's Medicare beneficiaries could be assigned to any in-scope primary care source. Patients went unassigned because: they received no reimbursed services; they made no ambulatory visits, they made visits only to specialists or undetermined physicians only; or their PCS was out-of-scope.

For the analysis reported here, we included only practices providing services to one or more patients with the in-scope conditions. Across the three states, this amounted to about 3,000 practices each treating on average, about 35 diabetics and 24 CHF patients.

The final stage of database development involved merging information from other sources that described the physician or group practice. We obtained provider files from the Medicare contracting carriers in the three states. These files were the files they used in their ongoing interaction with each physician.

These files contained information on the zip code practice location of each physician's main practice. We translated this zip code into county designations which we used to determine if the practice was located in a Metropolitan Statistical Area (i.e., urban) or a rural county.

We considered the physician to be practicing in the self-designated specialty in which they spent the majority of their time. Actual board certification information was not available consistently in all three states, but an analysis in Maryland suggested that approximately 70% of the physicians were certified in their self-professed specialty.

By analyzing the characteristics of all physicians who submitted bills jointly as a group practice, we were able to develop an algorithm to identify a physician's group practice affiliations. For the purpose of this study, we included only those group practices that were single specialty or mixed specialty primary care practices (i.e., family practice and internal medicine) or where primary care physicians represented the majority within a multispecialty group.

Study Variables and Measures

As part of the DEMPAQ study a great deal of energy was spent identifying services that were reliably documented in the claims database that could be used as indicators of quality of care. A set of recommended and limited-use procedures/services were chosen to be evaluated for both conditions reported here. Based on clinical guidelines, a recommended procedure was considered as a general standard of care for all patients, while a limited use procedure was characterized as a service that was indicated only in certain circumstances. The patients' receipt of the following procedures during the study period served as the main focus of the analysis:

Diabetes

Recommended Procedures
- HgbA1c (glycosolated hemoglobin)
- Ophthalmic Exam
- Total Cholesterol Level

Limited Use Procedures
- Blood Glucose

Congestive Heart Failure Recommended Procedures

- Serum Potassium

Limited Use Procedures

- Chest X-Ray
- ECG
- Echocardiogram

A further discussion and technical description of how each procedure was defined is provided in Appendix 1.

Specifically, the dependent measure used in the analysis is the proportion of in-scope patients within a primary care practice who received the procedure in question at least once during the 12-month study period. Although the unit of analysis was the primary care source practice, the service could have been delivered by any provider (e.g,, the PCS, an independent lab, a specialist, another primary care physician, or a hospital outpatient department).

The independent variables included practice attributes as follows:

- *Type of Practice (e.g., Specialty, Group or Solo):* This attribute is denoted by several binary measures. These variables are developed from the carrier-specific provider files used in the DEMPAQ analysis.

- *Location of Practice:* Metropolitan and rural practices are described as dummy variables. This distinction is based on whether the practice's location was in a Metropolitan Statistical Area (MSA). The second location variable is state.

- *Patient Age and Sex:* Based on the Medicare beneficiary file, these represented the proportion of patients with each condition in age and gender categories.

- *Patient Comorbidity:* Using diagnostic information provided by physicians on their claims during the year study period, we determined the number of unique types of conditions the patients were under treatment for. The categories are based on the Johns Hopkins Ambulatory Care Group (ACG) case-mix classification system (Weiner et al. 1991). Without regard to organ system, the ACG methodology groups clinically similar conditions (e.g., self-limited acute, or unstable chronic) into the same category. Thus, they represent broad clusters of conditions, the sum of which can be considered representative of an individual patient's burden of illness. Although the number of conditions included in the model is not a direct measure of the severity of the patient's diabetes or CHF, it does provide a measure of the case-mix of each physician's patient load. (Note: the number of conditions are, however, an indirect measure of severity given that most of the complications of diabetes or CHF [e.g., ketoacidosis, peripheral vascular disease, valve disease] are categorized into condition clusters that are separate from the primary condition.) The proportion of in-scope patients within the practice in each category was entered into the model as a measure of the practice's case-mix.

- *Practice Size and Case Load:* Using information from the NCH files we were able to categorize practices in terms of the number of overall (non condition specific) Medicare patients it served. Based on the distribution of practices in each state, each PCS was categorized into thirds and was designated as a high, medium or low size practice. We also developed a proportional measure of the condition-specific case load that represented the percentage of the practice's assigned Medicare patients who were diabetics or CHF patients.

Statistical and Analytic Approaches

The objective of our statistical analysis was to determine which physician practice or patient case-mix factors are associated with variation in proportion of diabetic and CHF patients receiving the procedure of interest. Each practice's condition-specific experience with a procedure was treated as a unique observation. For example, Practice A sees diabetic and CHF patients. Practice A's utilization of each condition's procedure of interest (eight in total) represent unique observations in one of the eight separate multivariate analyses reported here.

We used a linear probability regression (OLS) model to estimate the relationships between practice characteristics and procedure use. The distributions of the dependent variables tended to be fairly normal. In situations where the distribution of procedure use was less gaussian in nature, we tested the use of an arc-sine transformation. An arc-sine transformation takes data with an inverted S-shaped probability density function and makes them more normal. We found the OLS estimates to be more efficient than those obtained using the arc-sine transformation and thus chose to adopt this approach.

In situations where more than one dummy variable was used to characterize independent variables (e.g., three dummy variables were used to represent the states), one dummy variable was removed and used as a reference category to avoid colinearity. Since both the age and comorbidity categories were, in effect, linear combinations of each other, one of the age and comorbidity categories were thus removed and assigned as reference categories.

One of the analytic problems we encountered was "left-censored data." Left-censored data represent observations of the dependent measure that are missing or assigned a value of zero on the left hand side (the dependent variable side) of the regression equation. For example, if a physician had 15 diabetic patients, and none received a HgbA1c during the year, the physician's dependent measure for that regression would be zero, even though a full set of independent variables were available to describe the doctor's practice attributes.

We performed a sensitivity analysis for HgbA1C (where the problem was the greatest) to assess the impact of this censoring problem on interpreting the standard OLS regression to be reported here. To assess if our OLS models were truly significant, and not attributable to the bias of observations clustered at 0, we completed TOBIT analyses which explicitly considers the left-censoring problem by using MLE estimation. A comparison of the results of the two analyses indicated that the approximate magnitude and signs of all of the parameters were the same using both the TOBIT and OLS models, except for the practice size independent variable. The TOBIT results inverted the signs of the large and small total practice size variables. However, it should be noted that the practice size OLS parameter estimates for HgbA1C were not statistically significant.

Results

Table 1 (tables are located in Appendix 2) presents descriptive analyses of the approximately 168,000 patients included in this study by condition and state. This table presents the age/gender and case-mix characteristics, as well as number of practices serving each patient cohort. Note that although the number of patients included in the survey from Maryland is not greater than the other states, there are more practices serving as PCSs in that state, resulting in a smaller average case load in Maryland.

Table 2 presents the characteristics of the approximately 3,000 physician practices included in this analysis by state. This table presents the type of practice and specialty designation, location of practice and the overall (not condition specific) size of practice. Note that solo practices predominate in each state, general/family practice is more common in Alabama and Iowa than in Maryland, Iowa is the most rural of the three states, and Maryland is the least rural.

Table 3 presents a bivariate breakdown of the services provided to diabetics by characteristics of the practice. This table shows the average percentage of diabetics within a practice who did receive the service of interest. Note that the variation (or odds ratio) across the highest to lowest states for each service is as follows: HgbA1C -2.22; eye exams -1.35; cholesterol -1.33 and glucose 1.22%. (These ratios are all significant given the very large sample size.)

Table 4 presents a bivariate breakdown of the services provided to CHF patients by practice characteristic. The variation across the states for each service is as follows: potassium—1.16; x-ray—1.10; ECG—1.49; and echocardiogram—1.35.

Table 5 presents the linear regression results for diabetic services. Each regression was run using the physician practice as the unit of analysis. The parameter estimates on this table indicate the percentage change in the dependent measures (i.e., the proportion of patients in the practice receiving the service) associated with the independent variable, holding all else constant. In general, only those coefficients that are statistically significant should be considered. For example, for HgbA1C, this analysis suggests that holding all else constant, the patients in Maryland practices are 7.7% more likely to receive the procedure than those in Iowa (the reference category) and that patients in Alabama are 5.6% less likely to receive the procedure.

Some noteworthy results from the analyses on Table 5 are as follows:

- There does not appear to be a strong consistent impact of type of practice, although there is some indication that the patients of solo family/general practitioners and patients cared for by multispecialty groups receive fewer procedures than solo internists (the reference category).
- In general, patients of practices within metropolitan areas receive more services.
- It appears patient panels made up of older patients are less likely to receive the services.
- The greater patient panels' case-mix—as measured by more patients in higher illness burden categories—the more likely it is that the patient will receive the service.
- There is some suggestion that patients cared for by practices that are smaller than average are less likely to receive the services.
- Patients cared for by practices with a higher proportion of diabetics in their panel are more likely to have received the services. This is true for all procedures expect eye exams, where a higher percentage of diabetics is associated with a lower percentage of separate eye exams. (It is possible that fundus exams are more likely to be delivered by primary care physicians who specialize in diabetes and thus ophthalmologic consults are unnecessary.)
- The patterns across the states are not completely consistent, though the results suggest that Maryland patients are more likely to receive most services than patients in Iowa or Alabama. The Alabama results vary by procedure, with the lowest likelihood of receiving HgbA1C and eye exams, mid level use of cholesterol and highest use of glucose.

Table 6 presents the linear regression results for the four congestive heart failure procedures. The dependent measure was the percent of the CHF patients receiving the service over the year. The unit of analysis was the 2,926 physician practices across the three study states. Unlike the previous diabetes regression model, where three out of the four procedures of interest (all but blood glucose) were recommended, in the case of CHF, we considered only the serum potassium as a recommended procedure, with the other three being categorized as limited-use procedures.

Some noteworthy results from Table 6 are as follows:

- All else equal, patients of general/family practitioners are less likely to receive all four of the CHF related procedures—including recommended potassium—when compared to patients cared for by internists.
- Patients of urban physicians are much more likely to have received the recommended serum potassium test and somewhat more likely to have received the nonrecommended ECG. There is no statistical difference between urban and rural practices for the other two procedures.
- Physicians whose practices include older patients are less likely to perform or order all of the procedures. This inverse relationship is strongest for chest x-rays and echocardiography.
- Case-mix—as measured by the ACG comorbidity categories—has a very strong positive relationship with receipt of the tests.
- The size of the physician's overall practice does not appear to have a relationship with receipt of the procedures.
- Patients cared for by practices with a higher proportion of CHF patients are less likely to receive two of the nonrecommended tests (chest x-ray and echocardiography). Although not statistically significant, this measure of "specialization" was associated with a modest increase in the receipt of the potassium and ECG procedures.
- The patterns across the three states are fairly consistent. For all procedures, it appears that the physicians in Iowa order or perform significantly fewer tests. All other factors held equal, the practitioners in Alabama provide somewhat more potassium and x-rays, while those in Maryland use twice as many ECGs and echocardiograms.

Discussion

There is a burgeoning literature in the area of physician practice pattern variation. In part because of methodologic limitations, with few exceptions this past research has focused on inpatient hospital care and surgical services. To the best of our knowledge, the study reported here represents the first time that the primary care practices of the universe of office-based physicians within entire states have been assessed using small area variation technique. Moreover, by surmounting many technical challenges, this study represents one of the first broad applications of what managed care administrators and quality assurance professionals have come to call "quality profiling." Some implications of this study are discussed below.

This analysis indicates that existing claims data within the Medicare program can be used to portray the patterns of practice of office based practitioners to provide clinically meaningful and statistically valid comparisons that are relevant to quality improvement. Analyses such as those presented here could be used by physician groups as one source of input into the quality improvement process. This type of information should help increase the delivery of desired services and decrease the delivery of unnecessary services to patient populations.

This study indicates that, similar to inpatient and surgical services, clinical actions in the primary care sector exhibit fairly wide levels of practice pattern variation. This variation appears to be not only closely associated with patient related factors, such as case mix, but also to characteristics that are linked directly to the physician (e.g., specialty) and geography (e.g., location of practice).

For those clinical activities that are accepted as "ideal standards" of care, (e.g., HgbA1C or eye exams among diabetics) this study provides some evidence that quality improvement programs should be targeted more directly at certain types of physicians (e.g., general practitioners) in particular geographic areas (e.g., rural locales, or selected states). For other procedures where general wide-scale use is not deemed effective (e.g., blood glucose for diabetics or chest x-rays among CHF patients) this study indicates that certain provider and geographic factors are associated with potential overuse of the services. This aspect of the study underscores the need for better scientifically derived clinical practice guidelines and programs to implement them.

This study has developed methods that should have direct applicability for health services researchers and health plan analysts interested in further investigation of patterns of practice in the primary/ambulatory care arena. These innovations are described elsewhere in this report and include: the development of analytical files from raw health insurance claims; meaningful claims-derived indicators relevant to quality; methods to help control for patient case mix; and analytic approaches that emphasize consumer populations, patients and physicians, rather than nonclinically relevant billing transactions.

Over the next decade the U.S. will be moving toward a reformed health care system. All signs indicate that population-oriented managed care health plans that emphasize primary care will gain in importance. The methods developed for this study have direct relevance for the quality improvement programs that will be put in place by these plans. Furthermore, the findings reported here indicate that such concerted action will be critical, if the care delivered to Medicare beneficiaries and other Americans is to reach the highest achievable levels of quality.

Appendix 1

Appendix One: Specification of Procedures

Diabetes

Recommended Care	Data Source	Notes
Hemoglobin A1C Test	CPT: 83036	Since the hemoglobin A1C test gives a good indication of the level of glucose control, it is recommended every 2 to 3 months (ABFP, pg. 37) or every 6 months (ADA, p. 367).
Ophthalmology Visit	CPT: 92002-92019 92225-92260	A complete eye and visual examination should be performed annually (ADA, p. 367). This examination is important to detect early signs of retinopathy (Harrison's, p. 1753) or cataract (Branch, p. 784).
	90000-90060 90600-90699 90750,90760	These codes are acceptable if reported by an ophthalmologist. Optometry visits in an optical shop are probably not billed to Medicare.
Total Cholesterol	CPT: 82465,82470 83705	Total cholesterol should be tested annually (ADA, p. 367) because diabetics are prone to have arteriosclerosis and have an increased risk of myocardial infarction (Harrison's, p. 1733).

Limited Use Care	Data Source	Notes
Blood Glucose	CPT: 82947-82952	While many physicians use blood glucose to monitor patients with diabetes, guidelines from ABFP (p. 37) and ADA recommend hemoglobin A1C for diabetes monitoring. In some case, blood glucose test may be a valuable adjunct to the hemoglobin A1C test. The blood glucose test is listed in the limited use category because it is sometimes appropriate for patients taking insulin. However, Medicare data do not indicate which drugs are used by a patient and may not indicate whether diabetes is type I or type II. Home blood glucose monitoring is not captured by these data.

References:

Allerheiligen DA, Erwin DT, Galazka SS, Smith CW. Diabetes Mellitus Reference Guide, 3rd edition. American Board of Family Practice, 1990.

Branch WT. Office Practice of Medicine, 2nd edition, Philadelphia: WB Saunders, 1987.

Wilson JD, Braunwald E, Isselbacher KJ, et. al. Harrison's Principles of Internal Medicine. New York: McGraw-Hill, 1991.

Congestive Heart Failure

Recommended Care	Data Source	Notes
Serum Potassium	CPT: 80002-80019 80060,80062 80065,84132	Close monitoring os serum potassium is necessary for patients who are receiving diuretics, as most patients with CHF do (Cohn, 1989).

Limited Use Care	Data Source	Notes
Chest X-Ray	CPT: 71010-71035	The chest radiography provides an indication of underlying cardiac abnormality and information regarding the status of the failure itself (ABFP, CHF Guide, p. 11). While useful, the chest x-ray is not a very sensitive screening test for heart failure (Branch, p. 129). The history and signs detected at physician exam are more important for monitoring and managing the patient with CHF.
ECG	CPT: 93000-93014	While ECGs do not confirm the presence or absence of heart failure, they provide clues to the etiology of the failure (ABFP, CHF Guide, p. 11).
Echocardiography	CPT: 93307-93350	Echocardiography (sonography) can provide accurate measurements of cardiac chambers, wall thickness, and vascular structures (ABFP, CHF Guide, p. 12). This technique should be limited to patients suspected of having cardiac valve disease.

References:

Branch WT. Office Practice of Medicine, 2nd edition, Philadelphia: WB Saunders, 1987.

Cohn JN. Current therapy of the failing heart. Circulation 1988;78:1099-1107.

Nong A, Reams CJ. Chronic Heart Failure Reference Guide, 3rd edition. American Board of Family Practice, 1990.

Appendix 2

Table 1 - Characteristics of Patient Population by State & Condition

Patient Characteristics	Diabetes			Congestive Heart Failure		
	AL	IA	MD	AL	IA	MD
Number of Patients	37,479	29,913	29,996	24,323	29,958	17,215
Total Number of Practices Serving Patients	845	818	1317	836	808	1283
Males	34.4%	41.0%	39.8%	36.9%	38.8%	40.9%
Females	65.6%	59.0%	60.2%	63.1%	61.2%	59.1%
Aged 65-74 years	51.8%	53.6%	56.5%	32.4%	24.5%	33.5%
Aged 75-84 years	39.1%	40.9%	36.2%	44.9%	44.1%	43.8%
Aged 85 years and over	9.1%	12.7%	7.3%	22.7%	31.4%	22.7%
Burden of Patient Illness*						
- One Conditions	5.1%	4.7%	9.6%	3.4%	2.6%	5.4%
- 2-3 Conditions	24.5%	22.3%	33.5%	19.7%	17.1%	29.4%
- 4-5 Conditions	28.9%	26.7%	28.6%	27.9%	26.0%	29.5%
- 6-9 Conditions	33.7%	35.9%	24.1%	38.9%	40.1%	29.6%
- 10+ Conditions	7.8%	10.4%	4.1%	10.0%	14.2%	5.8%
Average Number of Patients per Practice	44.4	36.6	22.8	29.1	28.4	13.4

*Based on the ACG case-mix system.

Table 2 - Characteristics of Physician Practices

Physician Practice Characteristics	States		
	Alabama	Iowa	Maryland
PANEL SIZE	866	822	1329
TYPE OF PRACTICE			
Solo Practice:	74.1%	62.4%	76.9%
- General Practice	- 63.2%	- 72.3%	- 33.0%
- Internal Medicine	- 36.8%	- 27.7%	- 67.0%
Single Specialty Group Practice:	11.1%	19.7%	11.1%
- General Practice	- 45.8%	- 72.8%	- 32.4%
- Internal Medicine	- 54.2%	- 27.1%	- 67.6%
All Primary Care Group Practice:	4.3%	6.6%	2.7%
Majority Primary Care Group Practice:	10.5%	11.3%	9.3%
LOCATION OF PRACTICE			
Urban Practices	63.5%	43.7%	89.7%
Rural Practices	36.5%	56.3%	10.3%
Total Number of Medicare Beneficiaries	243,145	245,361	236,652
Average number of Patients peer Practice	285	299	178

Table 3 - Diabetes Related Care by Practice Characteristics: A Bivariate Analysis

Practice Characteristics	Proportion of Patients in Practice Receiving Services			
	Hemoglobin A1C	Eye Exam	Total Cholesterol	Blood Glucose
TYPE OF PRACTICE				
Solo Practice:	16.0%	45.2%	55.5%	80.7%
- General Practice	- 13.0%	- 43.8%	- 51.0%	- 81.3%
- Internal Medicine	- 18.9%	- 46.7%	- 60.1%	- 80.0%
Single Specialty Group Practice:	17.3%	48.1%	57.2%	83.0%
- General Practice	- 17.1%	- 46.3%	- 51.3%	- 81.0%
- Internal Medicine	- 17.4%	- 50.0%	- 63.7%	- 85.0%
All Primary Care Group Practice:	17.7%	47.7%	54.6%	80.9%
Majority Primary Care Group Practice:	17.0%	43.7%	50.2%	76.4%
LOCATION OF PRACTICE				
Metropolitan Setting	18.2%	46%	59.8%	82%
Non-Metropolitan Setting	12.0%	45.7%	44.6%	77.3%
Alabama	10%	38.3%	55.6%	89.2%
Iowa	15.1%	51.6%	45.7%	73.1%
Maryland	22.2%	47.2%	60.8%	79.7%

Table 4 - Congestive Heart Failure Related Care by Practice Characteristics: A Bivariate Analysis

Practice Characteristics	Proportion of Patients in Practice Receiving Services			
	Serum Potassium	Chest X-Ray	ECG	Echocardiography
TYPE OF PRACTICE				
Solo Practice:	64.4%	60.6%	60.0%	13.3%
- General Practice	- 61.4%	- 58.7%	- 53.4%	- 12.2%
- Internal Medicine	- 67.4%	- 62.5%	- 67.0%	- 14.4%
Single Specialty Group Practice:	69.6%	63.3%	57.9%	14.2%
- General Practice	- 67.5%	- 59.7%	- 50.9%	- 11.6%
- Internal Medicine	- 71.7%	- 67.3%	- 65.6%	- 17.0%
All Primary Care Group Practice:	66.5%	62.0%	53.2%	11.9%
Majority Primary Care Group Practice:	61.3%	65.2%	59.7%	14.0%
LOCATION OF PRACTICE				
Metropolitan Setting	69.0%	62.4%	63.5%	14.0%
Non-Metropolitan Setting	55.6%	59.5%	50.2%	12.4%
Alabama	68.1%	64.8%	56.6%	13.2%
Iowa	58.7%	58.9%	46.5%	11.2%
Maryland	66.6%	61.1%	69.4%	15.1%

Table 5 - Practice Level Regression Results for Diabetes

	Hemoglobin A1C	**Eye Exam**	**Total Cholesterol**	**Blood Glucose**
OVERALL MODEL RESULTS				
Model Adjusted R-square	0.0740	0.1566	0.1320	0.1000
Dependent Variable Mean	0.16321	0.45911	0.55175	0.80591
Intercept	0.084 a	0.383 c	0.384 c	0.640 c
PARAMETER ESTIMATES				
General Practice	-0.010	-0.028 c	-0.035 b	0.027 b
Osteopathic Medicine	0.012	-0.031	0.082	0.206 a
Majority PC Group	0.013	-0.060 c	-0.063 c	-0.031 a
All Primary Care Group	0.035	-0.022	0.007	0.029
Single Specialty Group	0.015	-0.010	0.028	0.025
Metropolitan Practices	0.027 a	0.008	0.096 c	0.042
Female Patients	-0.049	0.031	-0.020	0.029
Patients aged 75-84 years	-0.017	0.028	-0.100 b	-0.059 a
Patients aged >= 85 years	-0.047	-0.074 a	-0.124 a	-0.173 c
Patients With:				
No Condition	-0.091	-0.282	-0.233	-0.086
One Conditions	-0.237 c	-0.066	-0.316 c	-0.208 c
4-5 Conditions	0.069 a	0.176 c	0.033	0.026
6-9 Conditions	0.153 c	0.217 c	0.232 c	0.119 c
10+ Conditions	0.164 c	0.240 c	0.407 c	0.173 c
Large Total Practice Size	-0.019	0.027 c	-0.020	0.020
Small Total Practice Size	0.002	-0.029 c	-0.011	-0.041 c
Condition-specific Caseload	0.205 c	-0.119 c	0.114	0.220 c
Maryland Practices	0.077 c	-0.009	0.150 c	0.085 c
Alabama Practices	-0.056 c	-0.132 c	0.084 c	0.146 c

Significance: a=.05 b=.01 c=.001.

Reference categories: Rural practices. Internal medicine practices. Iowa. Male Patients. Patients aged 65-74 years. 2-3 medical conditions. Medium total practice size.

N=2,979

Table 6 - Practice Level Regression Results for Congestive Heart Failure

	Serum Potassium	Chest X-Ray	ECG	Echocardiography
OVERALL MODEL RESULTS				
Model Adjusted R-square	0.0999	0.1173	0.2155	0.0422
Dependent Variable Mean	0.648	0.61536	0.59439	0.13485
Intercept	0.411 c	0.589 c	0.411 c	0.131 c
PARAMETER ESTIMATES				
General Practice	-0.029 a	-0.055 c	-0.079 c	-0.023 c
Osteopathic Medicine	0.080	0.082	-0.139	0.179 b
Majority PC Group	-0.040 a	0.034 a	-0.025	0.004
All Primary Care Group	0.017	0.001	-0.062 b	-0.013
Single Specialty Group	0.046	0.032 b	0.0005	0.017
Metropolitan Practices	0.125 c	0.010	0.037 c	-0.008
Female Patients	0.043	-0.062 b	-0.028	-0.019
Patients aged 75-84 years	-0.020	-0.076 c	-0.028	-0.045 b
Patients aged >= 85 years	-0.004	-0.100 c	-0.057 a	-0.098 c
Patients With:				
No Condition	0.117	-0.297	-0.554 b	-0.093
One Conditions	-0.398 c	-0.242 c	-0.246 c	0.016
4-5 Conditions	0.032	0.125 c	0.048	0.053 b
6-9 Conditions	0.173 c	0.197 c	0.188 c	0.070 c
10+ Conditions	0.287 c	0.393 c	0.373 c	0.169 c
Large Total Practice Size	0.0149	-0.008	-0.003	0.002
Small Total Practice Size	-0.021	0.006	-0.003	0.017 a
Condition-specific Caseload	0.088	-0.408 c	0.080	-0.146 b
Maryland Practices	0.075 c	0.039 b	0.238 c	0.040 c
Alabama Practices	0.086 c	0.074 c	0.101 c	0.022 b

Significance: a=.05 b=.01 c=.001.
Reference categories: Rural practices. Internal medicine practices. Iowa. Male Patients. Patients aged 65-74 years. 2-3 medical conditions. Medium total practice size.
N=2,926

Chapter 20
Patterns of Variation, Record Review

Abstract

Background
This chapter presents data on the patterns of variation in ambulatory medical practice. First, data describing how practice varies by state, specialty and patient cohort are presented. The focus is on understanding the types of analyses possible with the data rather than identifying statistically significant findings. Second, the effect of documentation on performance scores is explored. Finally, this chapter presents some hypothetical sample size calculations. What size physician and patient sample is required to demonstrate specific differences?

Methods
For the descriptive variation analyses, record review data from each state were cross-tabulated, specialty by state. Additionally, cross-tabulations by patient group, e.g., by age range, or drug therapy, were conducted as appropriate. Some analyses were conducted by comparing performance across diagnoses, drugs or tests. As samples were small and the analyses exploratory, statistical significance levels were not calculated. The effect of documentation on performance scores was explored by calculated Spearman correlation coefficients. Variance component analysis permitted estimation of sample sizes.

Results
The data from the DEMPAQ pilot test permit a variety of comparisons. Although these results should be viewed with caution as they are derived from pilot test data, they demonstrate variability in practice and raise potentially important clinical and practice organization issues.

There is no clear link between record documentation and performance scores. The correlation is positive but weak. Better documentation does not necessarily ensure a higher performance score. Thus the variation illustrated in this chapter probably stems from differences in physician performance rather than differences in documentation.

Sample sizes (both physician and patient) required to demonstrate a 5 point improvement at 95% power, in drug or diagnosis monitoring, diagnosing performance, and preventive services are reasonable and most likely feasible in an environment of constrained resources.

Discussion
Data on variation should be used to initiate dialogue with practicing physicians about the findings, their interpretation and barriers to improvement. Repeated measurement over time will help benchmark levels of performance.

Background

Variation in medical practice typically provides fertile ground for health care analysts and researchers. Wennberg's landmark approach to describing variation (small area analysis) marked initiated debate about the reasons for the variation which continues to this day. This chapter presents some initial investigations into practice variation in three states and between three primary care specialties. It also explores the possible effect of documentation on performance scores. Do better scores result only from better documentation? Finally, the chapter presents some sample size calculations.

Rather than being organized by the traditional methods, results and discussion, this chapter is ordered by type of variation study: patterns of variation, effect of documentation on performance scores, and sample size calculations.

Patterns of Variation

Methods

Record review data from each state were contained in six files: drugs; diagnoses; tests; procedures; prevention; and visits. With the exception of the visits files, the files corresponded to the DEMPAQ functions. The visit file combined follow-up and record documentation data.

Using SAS, data were cross tabulated by state and specialty for each diagnosis, each test, selected drugs, prevention and follow-up. Additional tabulations were done for diabetes (by drug therapy), for hypertension (by drug therapy), for osteoarthritis (joint pain complaint), for CHF and COPD (dyspnea complaint). These data form the basis for a series of graphic data displays, prepared in Harvard Graphics.

Results

The following charts illustrate potentially interesting findings and demonstrate how the data can be used to initiate dialogue with physicians about the findings. Each graph is accompanied by text:

- explaining to physicians how the graph may be read;
- listing known *caveats;*
- posing questions that the results raise.

The text for each chart immediately precedes the graph.

Evidence for a New Diagnosis,Three State Comparison

The chart on the next page shows the percent acceptable performance for making a new diagnosis based on specific DEMPAQ criteria. The number of patients with a new diagnosis are shown underneath the diagnosis abbreviation, along the x axis. The y axis displays the percent acceptable performance based on the above criteria. Nurse reviewers abstracting the office record looked for the information identified in Table 1 as evidence supporting a new diagnosis of the listed condition.

Table 1
Evidence for a New Diagnosis

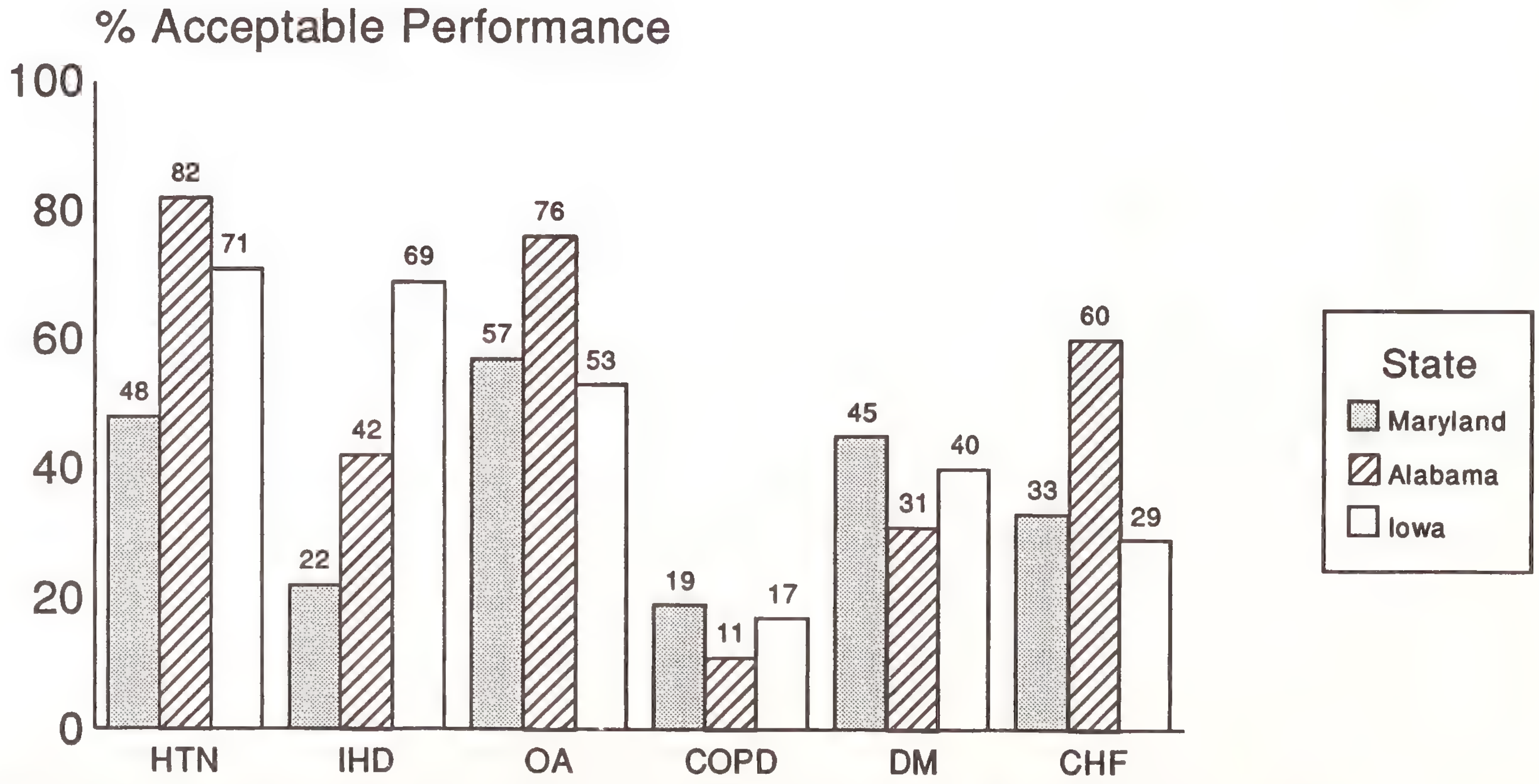
Condition	Criteria indicating appropriate evidence for a new diagnosis is found in the medical record
Ischemic Heart Disease	Cardiac chest pain OR exercise tolerance test OR positive cardiac catheterization
Diabetes	Non-fasting glucose level > 250 with symptoms of DM OR Fasting glucose level > 140 on two occasions OR Oral glucose tolerance test > 150 at 2 hrs and some point between 0 and 2 hrs
Congestive Heart Failure	S3, pulmonary rates, raised JVP and/or pedal edema, cardiac enlargement OR left ventricular ejection fraction <40% OR signs of CHF on Chest xray
Osteoarthritis	Symptoms for longer than 1 month that are suggestive of osteoarthritis OR positive xray findings of osteoarthritis
Hypertension	Diastolic blood pressure >= 90 on two occasions OR systolic blood pressure >= 140 on two occasions OR any diastolic blood pressure >=130
Chronic Obstructive Pulmonary Disease	Productive cough lasting 3 months for 2 consecutive years with dyspnea on exertion OR FEV1 < 60%

There are a variety of possible explanations for the lower performance rates for COPD. The data may include patients for whom the physician was listing a "rule-out" diagnosis. This is unlikely since "rule-out" diagnoses were explicitly excluded from the review. Perhaps the criteria are invalid as peer review of the criteria showed varying levels of agreement with the criteria: 43% hypertension; 24% diabetes; 63% COPD; 65% CHF; 39% osteoarthritis; 59% IHD (see also Chapter 14). Another possible explanation for the data is that physicians may label patients as having COPD without documenting the appropriate evidence for the diagnoses in the medical record.

- Is there a possibility that this condition is over diagnosed?

Evidence for a New Diagnosis

Percent Acceptable Performance, By Diagnosis



	HTN	IHD	OA	COPD	DM	CHF
Maryland n=	56	32	141	48	31	52
Alabama n=	55	36	127	46	26	52
Iowa n=	24	16	74	18	10	28

Diabetes Monitoring by Type of Therapy

The next three graphs concern treatment of diabetic patients. The first two graphs show compliance with diabetes monitoring recommendations, by type of therapy and by physician specialty. The monitoring criteria and how often the monitoring actions should be performed are shown on the x axis. The percent of records which demonstrated compliance with the criteria is shown on the y axis.

The first graph compares the performance on the monitoring criteria by type of therapy: diet alone, oral hypoglycemic and insulin. It appears there may be differences in monitoring, by type of therapy. That is, patients on diet only therapy are the least likely to have monitoring. Patients on insulin are the most likely.

- Is there a clinical reason for the difference in monitoring compliance, i.e., does the risk to the patient of having diabetes vary by the type of therapy?

Keeping in mind that the ADA came out with its recommendation to perform regular HgbA1C testing in 1989 and that the period covered by the DEMPAQ chart review is 1989–1991:

- Why does it appear as though diabetic patients are not receiving HgbA1C tests?
- What is the role of HgbA1C in monitoring diabetes; how important is it?
- What are the barriers to receiving a regular HgbA1C test?

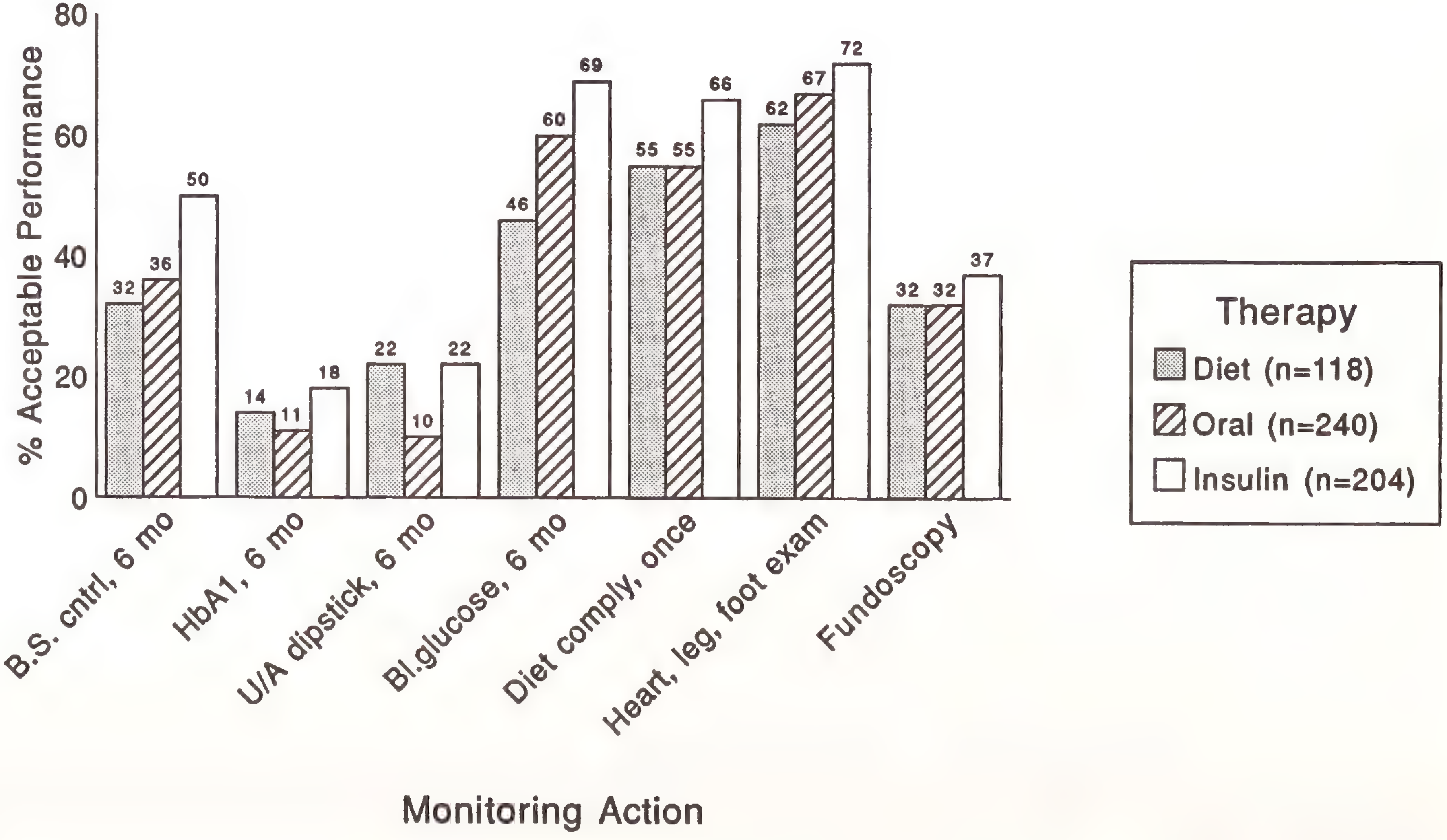
The charts reviewed often lacked a notation that a U/A dipstick for protein had been done; and a comment on blood sugar control.

- How important is an annual U/A dipstick for protein?
- Is it possible that the test was done but not documented?
- How likely is it that the physician inquired about blood sugar control but failed to document the response? Is this something that is important to document?

Fundoscopy appears to either not be performed or not be documented.

- What is the responsibility of the primary care physician with respect to making sure diabetic patients receive fundoscopy?

Diabetes Monitoring: By Therapy
Three State Combined Data

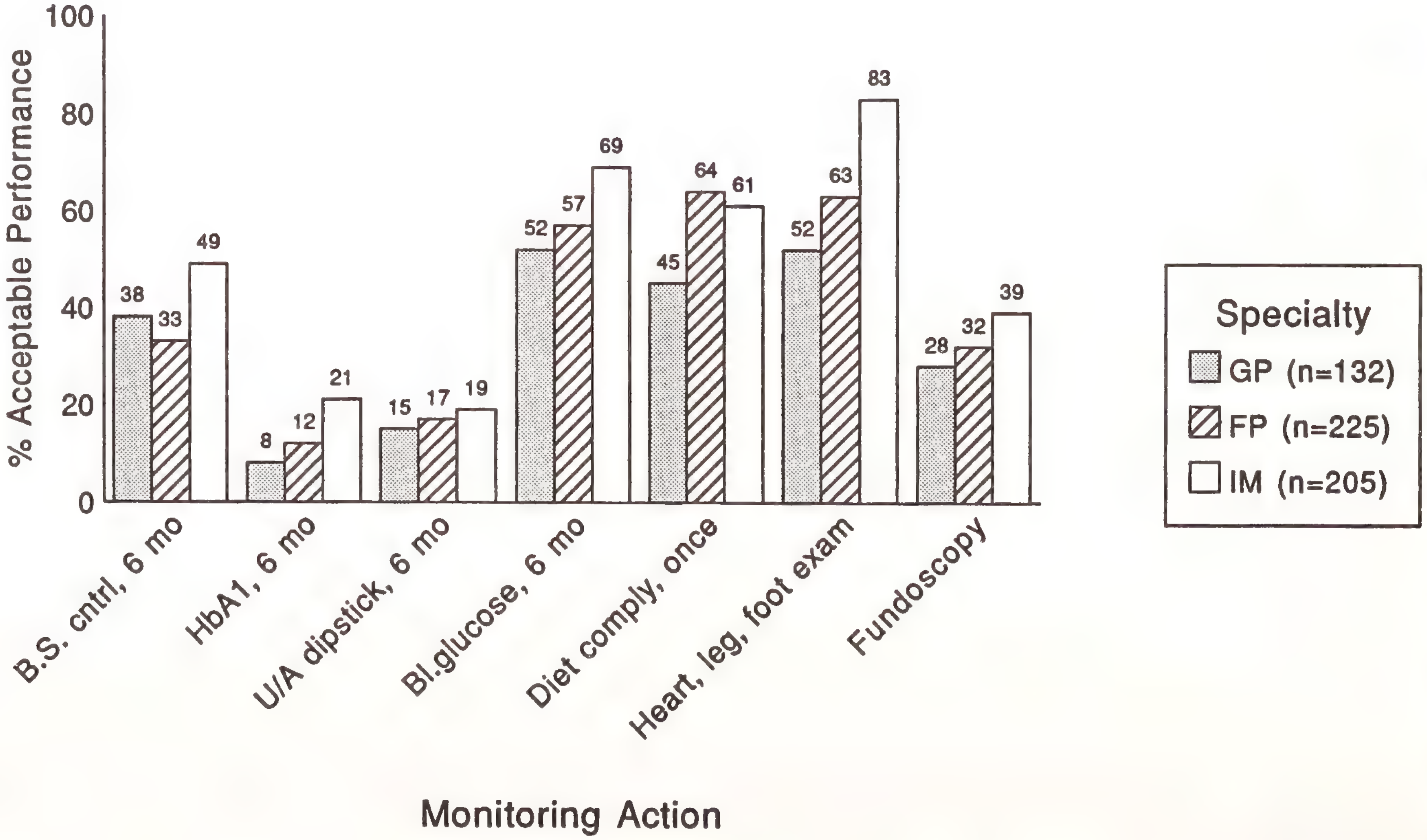


Diabetes Monitoring by Specialty
The next graph compares monitoring by specialty of the physician. The number of patients the graph is based on is shown in the legend. The same trends noted earlier are apparent.

In addition, there may be specialty differences in performing recommended monitoring actions with internists being most likely to perform the monitoring and general practitioners, least likely.

Diabetes Monitoring: By Specialty

Three State Combined Data



Insulin Monitoring

The final graph in the series of diabetes graphs shows a three state comparison of the monitoring actions that are recommended for patients taking insulin. The left hand side of the graph displays data for patients who were taking insulin prior to the beginning of the two-year review window. The right hand side of the graph shows monitoring for patients who began insulin during the two-year review period.

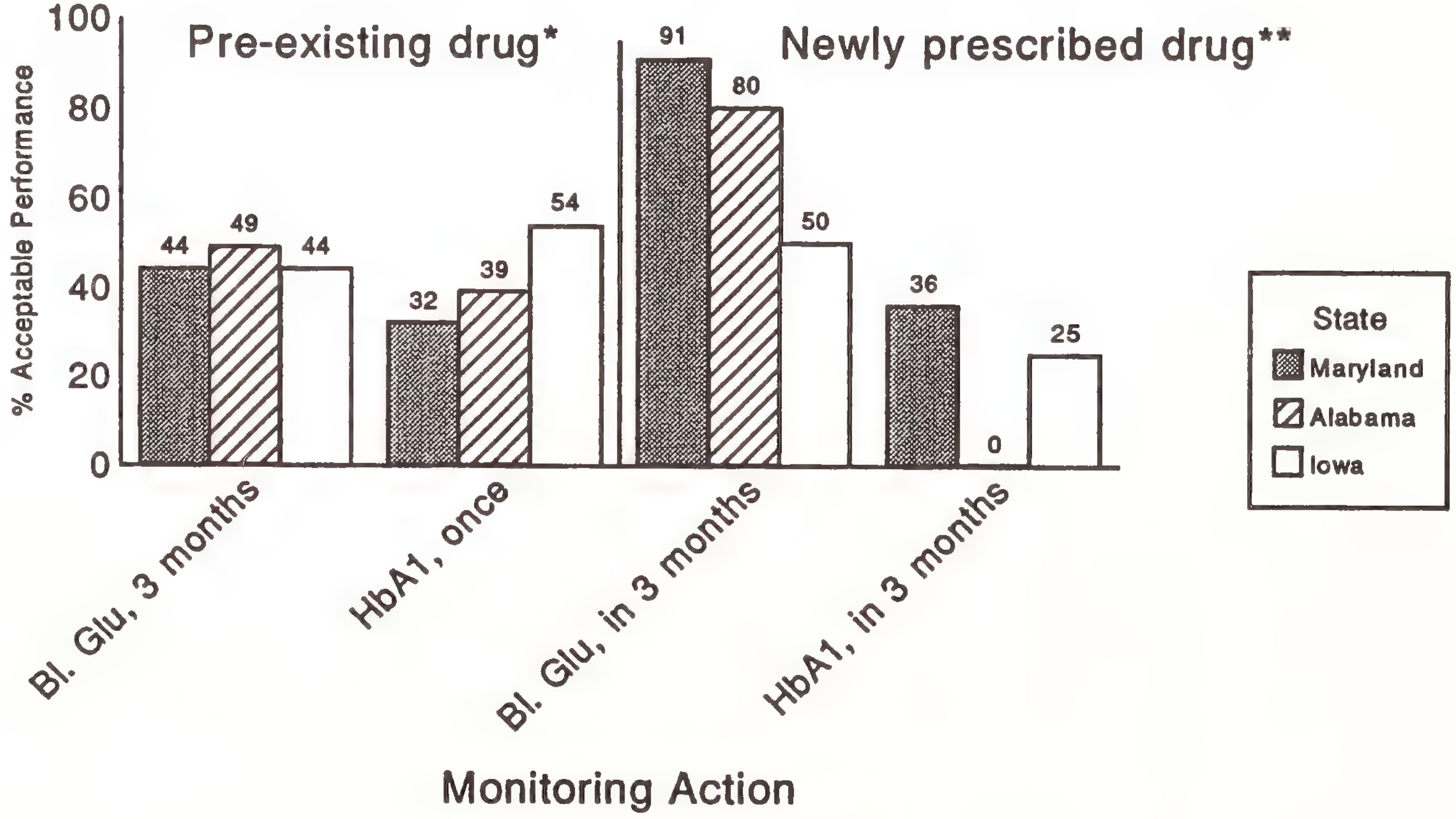
"Preexisting" includes patients who have been on insulin most of their lives and those who were started on insulin recently. Preexisting also includes well controlled and not so well-controlled diabetics.

This graphs seems to confirm the finding shown in the first graph: diabetics on insulin do not receive regular HgbA1C tests.

- Should patients on insulin receive regular HgbA1C tests?

Insulin Monitoring

Three state comparison



Sample sizes	Maryland	Alabama	Iowa
* Pre-existing drugs	57	86	39
**Newly prescribed drugs	11	5	8

Hypertension Monitoring

The graph on the next page presents data for patients with hypertension. The monitoring is shown by type of drug therapy. The drug therapies listed represent ones commonly prescribed by physicians for the treatment of hypertension. At the time of the criteria development, any one of the four drugs represented appropriate first choice drug therapy for patients with hypertension. The most recent Joint National Committee report (*Arch Intern Med*, 1993; 153:154-183) recommends only beta blockers and diuretics as first line antihypertensive drugs.

The graph shows that the monitoring of hypertensive patients varies little by type of drug therapy. Although it may appear that almost half of the patients with hypertension do not get their blood pressure checked, the percentage probably reflects the absence of a pulse measurement rather than the absence of a BP.

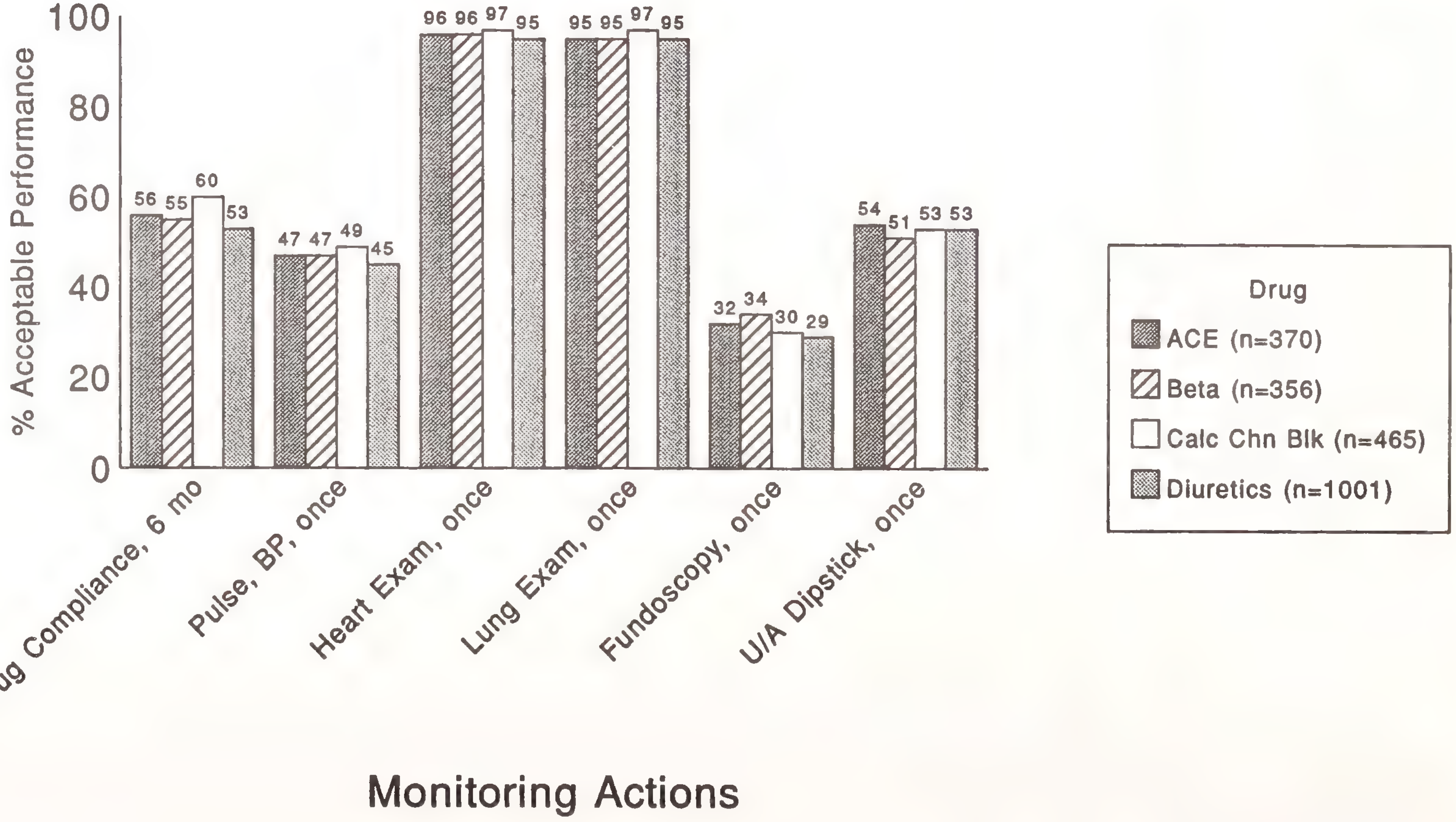
The data suggest that fundoscopy and U/A dipstick for protein are not routinely performed for patients with hypertension.

- How important are routine fundoscopy and U/A dipstick for elderly patients with hypertension?
- If these services are important, what can be done to remind the physician when these services are due?

Record Review Results: Hypertension

Monitoring by Type of Drug Therapy

Three State Combined Data



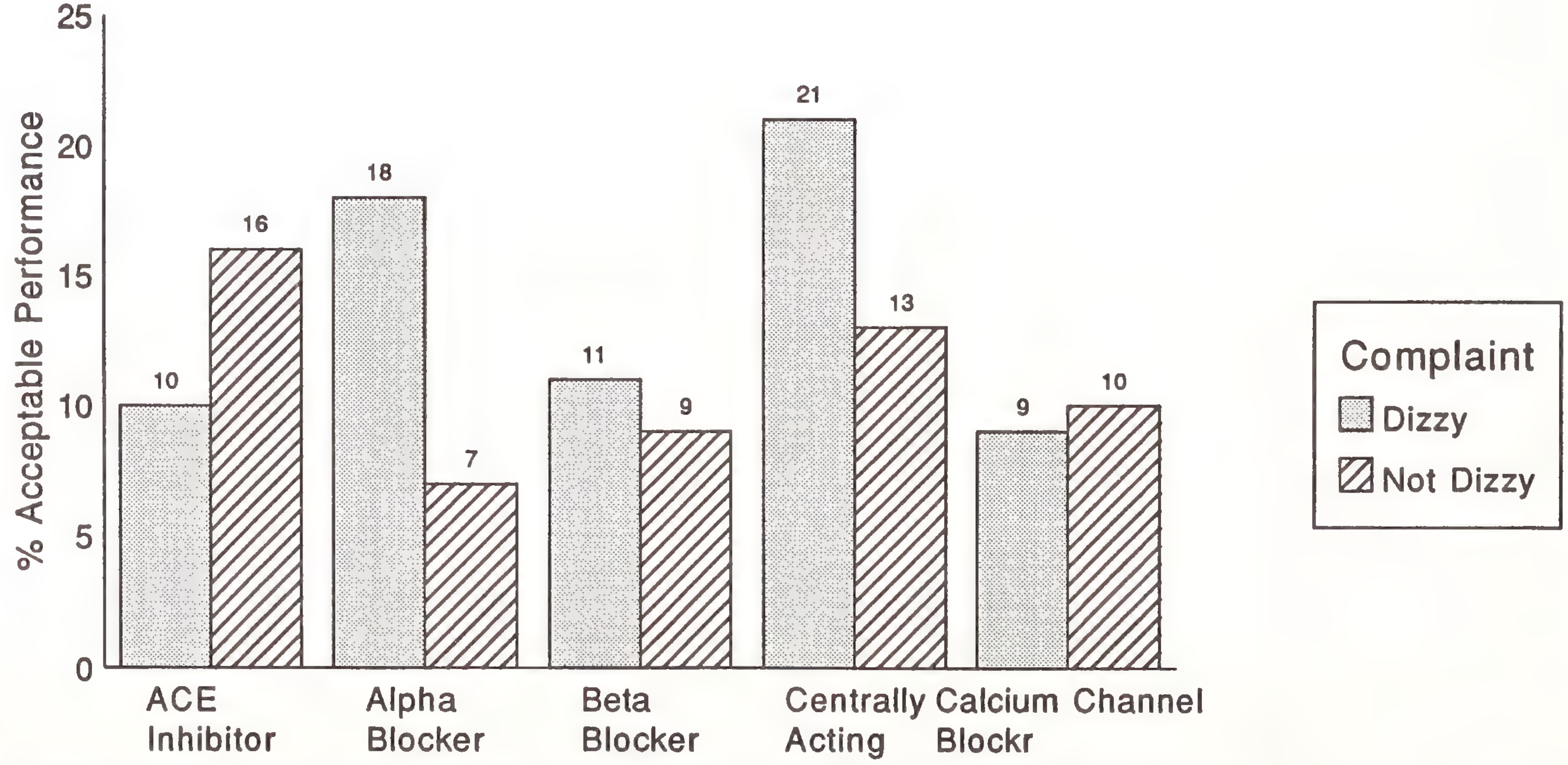
Standing Blood Pressure

The DEMPAQ criteria recommend annual standing blood pressure for patients taking certain drugs. The graph on the next page shows compliance with this criteria depending on whether the patient had complained of dizziness during the two-year review period. Although dizzy patients are slightly more likely to having their standing blood pressure measured, overall, few patients receive standing blood pressure measurements.

- Should routine standing blood pressure measurements be performed for patients taking certain anti-hypertensive drugs?
- How can the office routine be organized to get regular blood pressure (both standing and nonstanding) measurements?

Record Review Results

Standing BP for Patients on Anti-Hypertensive Drugs
Three State Combined Data



Dizzy n=	73	11	89	28	127
Not Dizzy n=	159	29	236	128	268

ACE Inhibitor Monitoring

The following graph shows monitoring of patients taking ACE inhibitors, by specialty. Monitoring actions and their time frames are displayed along the x axis. Two monitoring actions for preexisting drugs (that is, drugs started before the beginning of the two-year review period) are shown on the left. Four monitoring actions for newly prescribed drugs (prescriptions issued after the review began) are shown on the right.

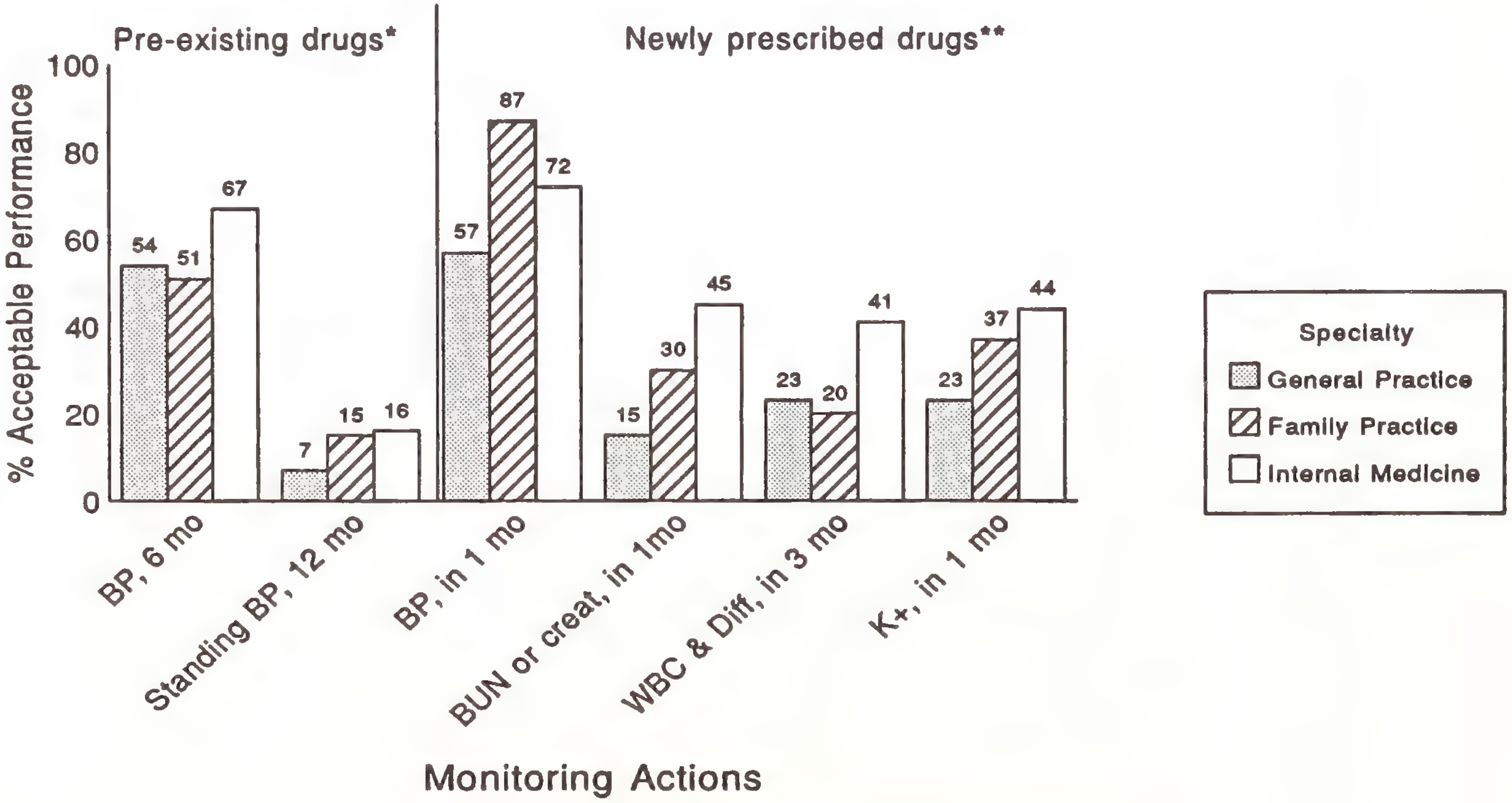
It appears there are some differences in monitoring, depending upon the specialty of the primary care physician. Internists appear more likely to perform most monitoring actions; family physicians to monitor the blood pressure at the start of a course of treatment. Standing blood pressures are not taken, regardless of specialties. Physicians appear to order chemistries for patients with newly prescribed drugs less than 50% of the time.

- How common is it for patients to experience a decline in renal function while taking ACE inhibitors, and therefore how important to measure BUN or creatinine after initiating treatment?
- Is neutropenia a rare or common problem with ACE inhibitors, and thus how important is it to perform a WBC and diff on new patients?
- Would "Post-It" type notes or flow sheets help remind physicians when these tests should be ordered?

Ace Inhibitor Monitoring

By Specialty

Three State Combined Data



Sample Sizes	GP	FP	IM
* Pre-existing drugs	57	92	138
** Newly prescribed drugs	47	86	86

Calcium Channel Blocker Monitoring

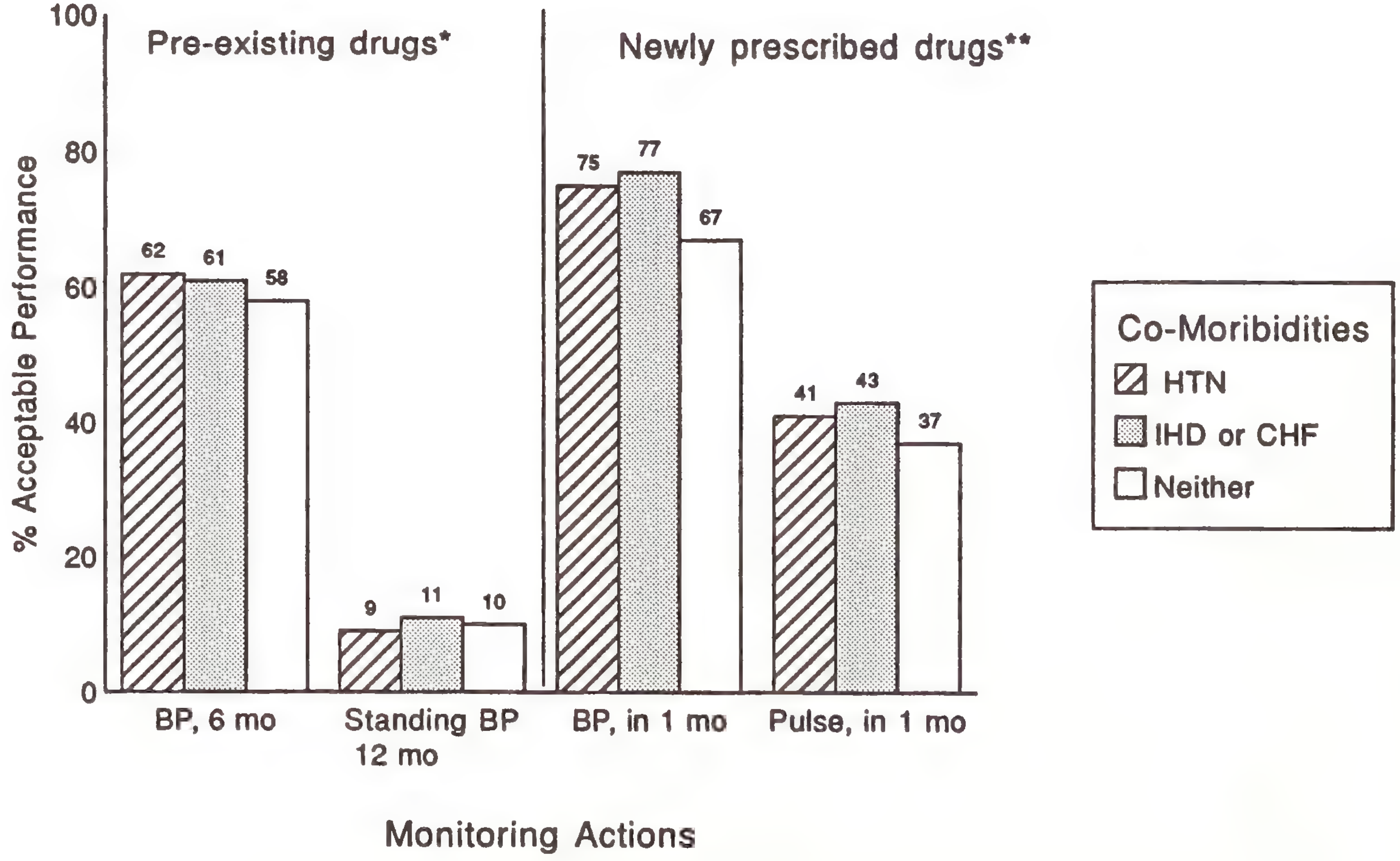
The following graph shows monitoring of patients taking calcium channel blockers by selected conditions. As with the previous graph, monitoring actions and their time frames are displayed along the x axis.

There are small differences in the monitoring compliance, depending on the patient's condition. Standing blood pressures tend not to be performed and pulse measurements, if done, are not likely to be recorded.

- Is it important for patients on new prescriptions of calcium channel blockers to have their pulse checked within one month of beginning therapy?
- How important is it for patients to have their blood pressure measured within one month of initiating treatment?

Record Review Results: Calcium Channel Blocker

with Selected Co-Morbidities



Sample Sizes	HTN	IHD or CHF	Neither
* Pre-existing drugs	271	202	158
** Newly prescribed drugs	179	64	52

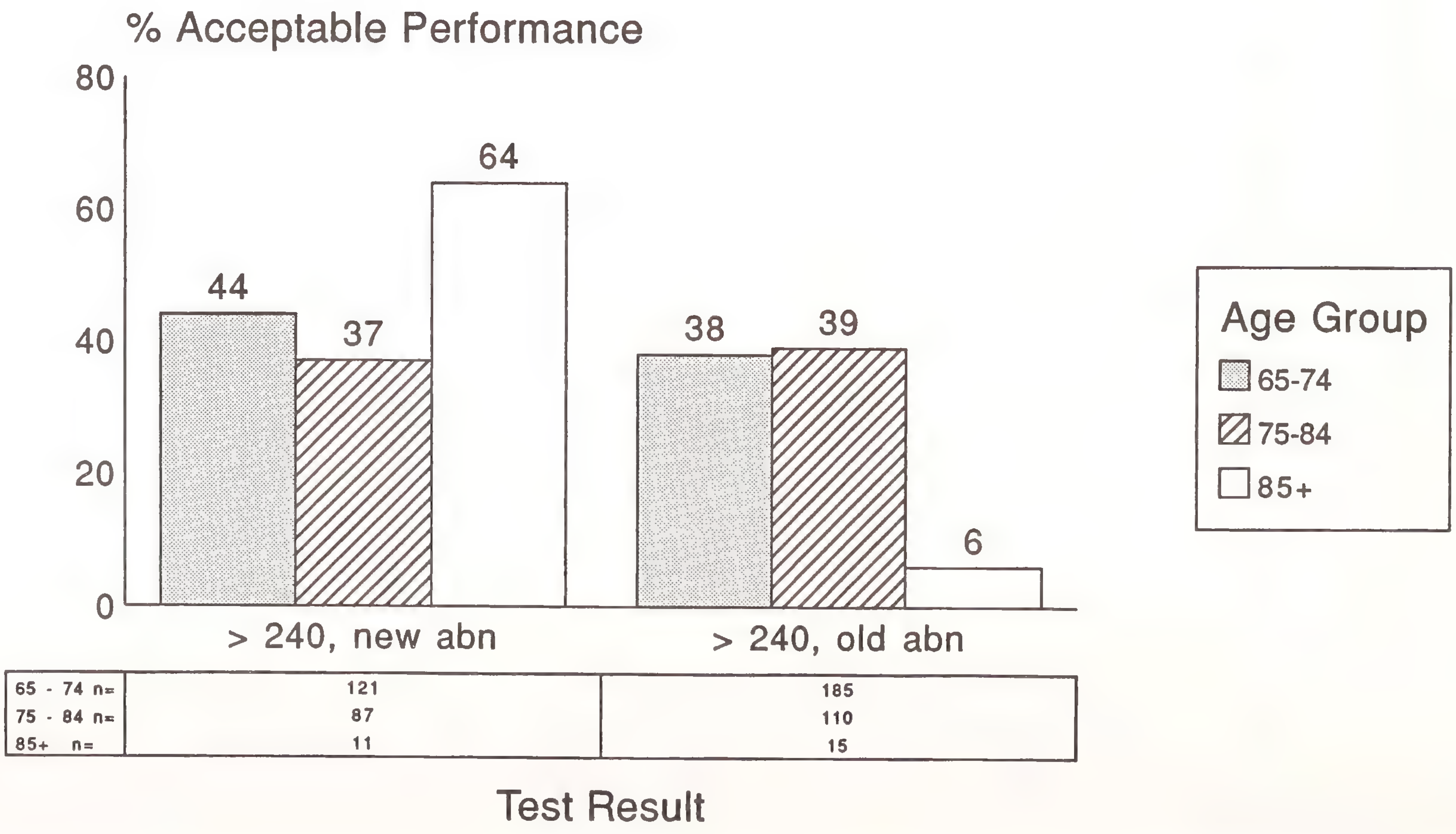
Cholesterol

The following chart shows follow-up of abnormal cholesterol results. On the left side of the graph, follow-up of a new abnormality is shown for three age groups. The right side displays the results of follow-up of preexisting abnormalities.

This graph apparently shows that physicians are most likely to follow-up new abnormalities in patients who are age 85 and older. Caution is urged in interpreting these results however since the number of cases in the 85+ age group is small.

Record Review Results: Cholesterol

Three State Combined Data



	> 240, new abn	> 240, old abn
65 - 74 n=	121	185
75 - 84 n=	87	110
85+ n=	11	15

Record Documentation

The ability to interpret the data gathered by the DEMPAQ project depends, in part, on the quality of the documentation in the medical record. Since the sample of records reviewed by the project came from volunteer participants, one might expect the quality of documentation to be better than average.

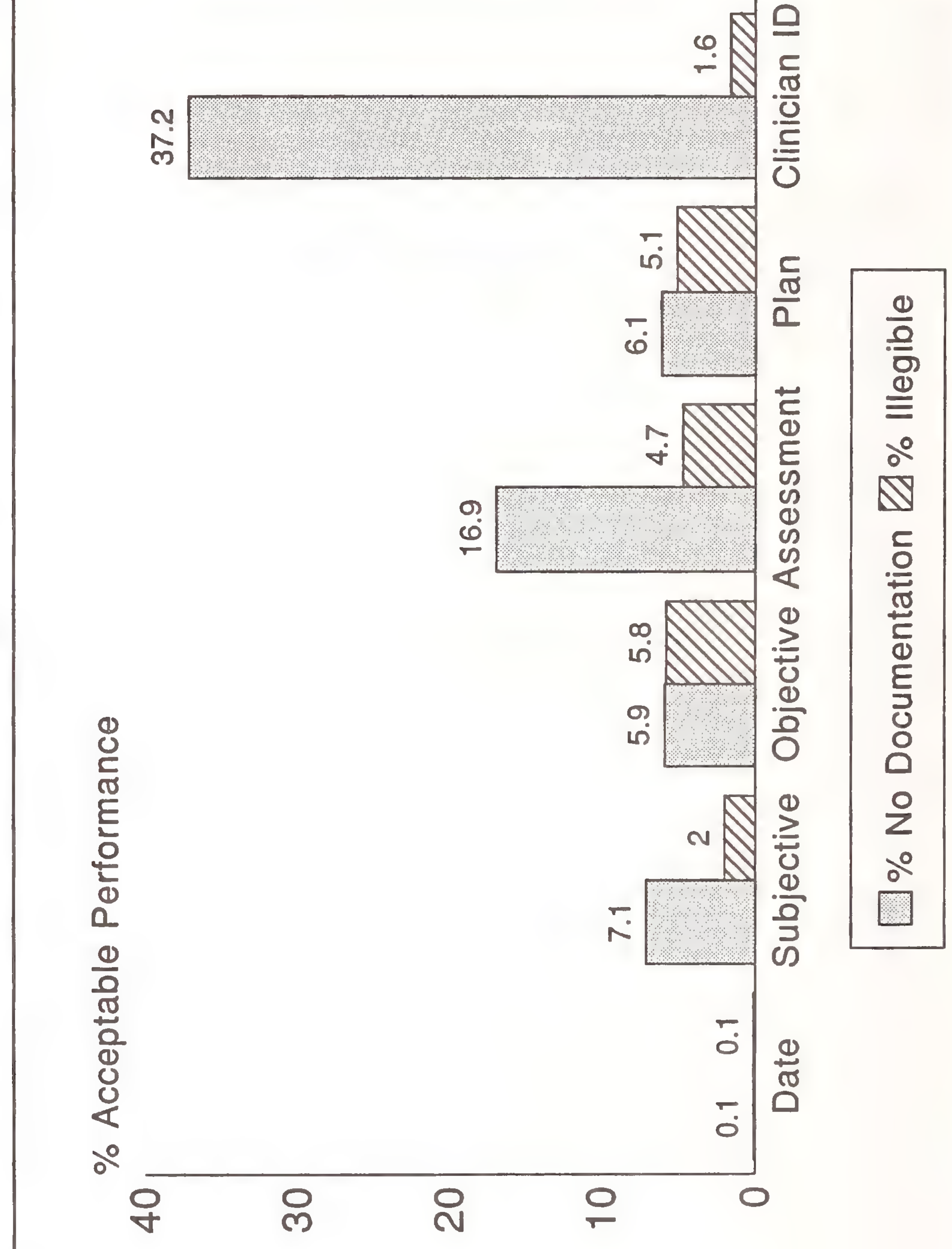
The graph on the next page shows the percentage of visits reviewed that had no documentation or illegible documentation in each of six categories. Virtually all the visits reviewed included a date. Approximately 9% of the visits lacked a readable notation of the patient's presenting complaint, 12% needed physical findings, 21% did not have a physician's assessment of the situation, 11% lacked a plan and almost 40% did not have a clinician signature.

Comparing these data across study states (next graph) suggest that there is some variation by state in the level of documentation.

- How important is legible and complete documentation?
- How can the task of documentation be simplified so that the physician has more time for direct patient care activities?

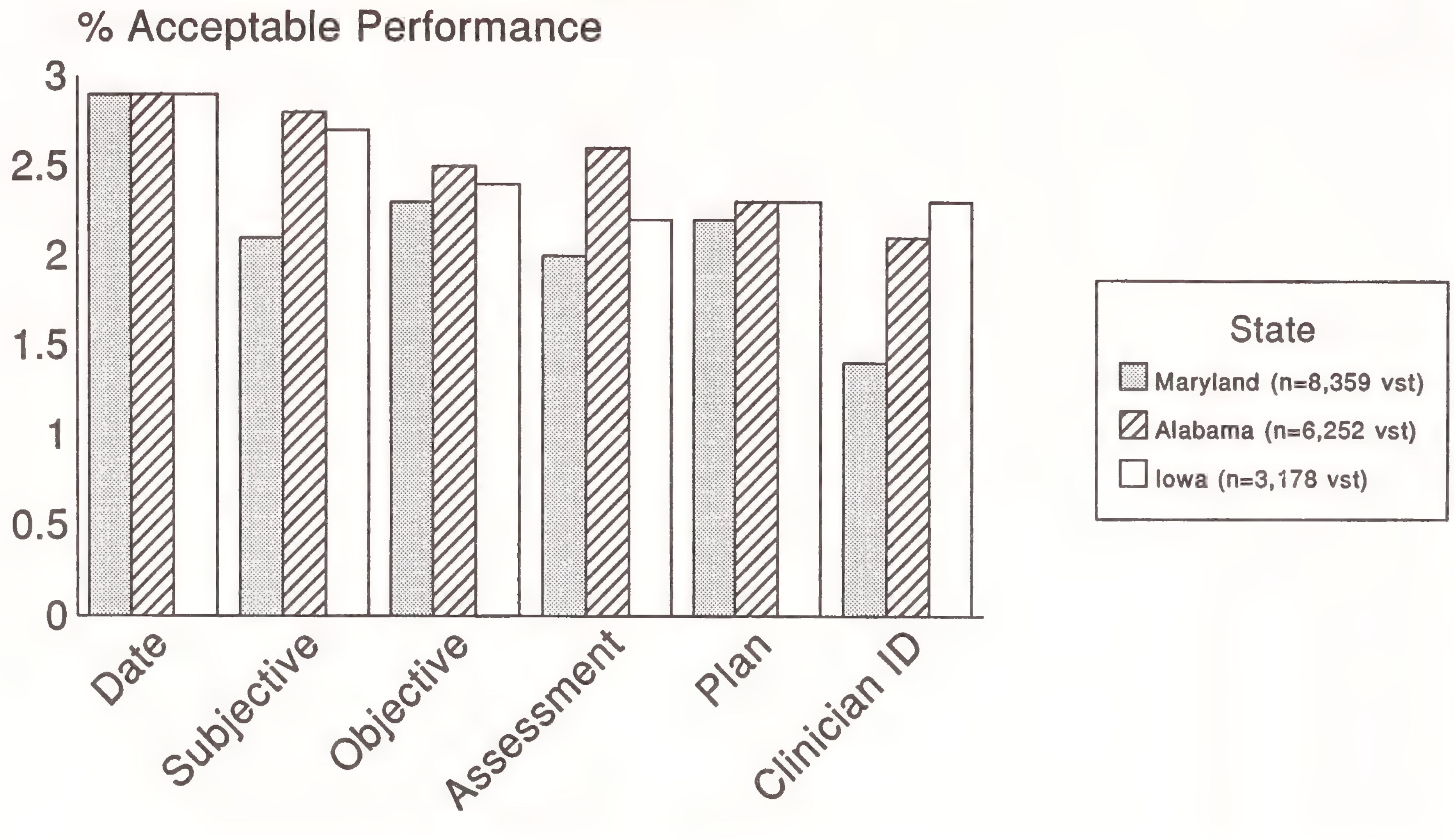
Record Documentation Results

Three State Combined Data



Record Documentation Assessment

A Three State Comparison



Documentation in the Record and the Effect on Performance Scores

One concern about performance scores derived from office record data is that better documentation will automatically result in better performance scores. If this were true then a high performance score would mean only that the physician was better at documentation, not that the quality of care was superior. To exam this possible association, we calculated Spearman correlation coefficients for the record documentation scores versus the performance scores for each function.

The following tables display, by function, the correlation coefficients. Only significant (at the .05 level) correlations are shown. Most of the correlations shown are highly significant at the less than .0001 level.

Table 2
Spearman Correlation Coefficients:
Drug Scores versus Record Documentation
For All 3 States and 3 Specialties Combined

Drugs

Documentation Category	Indicator				
	Dose and Frequency	Drug Indications	Drug-condition Contra-indications	Drug-drug contra-indications	Drug monitoring
Date	.11	.07	-.06	-.09	-.11
Subjective	.10	.01	.13+	.13*	.45*
Objective	.26*	.28*	.07	.08	.45*
Assessment	.24*	.23*	.05	.16*	.40*
Plan	.36*	.27*	.03	.06	.45*
Clinician ID	.16*	.15+	.07	.14+	.38*
Overall score	.30*	.26*	.05	.14+	.50*

* $p < 0.01$
+ $p < 0.05$

Table 3
Spearman Correlation Coefficients:
Test Scores versus Record Documentation
For All 3 States and 3 Specialties Combined

Tests

Documentation Category	Indicator		
	Test Indications	Urgent Response	Appropriate Response
Date	.06	-	-.02
Subjective	.33*	.03	.43*
Objective	.27*	.17	.31*
Assessment	.39*	.10	.36*
Plan	.19*	.08	.34*
Clinician ID	.22*	-.06	.33*
Overall score	.31*	.02	.43*

* $p < 0.01$
+ $p < 0.05$

Table 4
Spearman Correlation Coefficients: Diagnosis Scores versus Record Documentation For All 3 States and 3 Specialties Combined

Diagnosis

Documentation Category	Indicator		
	Old diagnosis monitoring	Evidence for a new diagnosis	Search for treatable causes/co-morbidities
Date	-.04	.10	-.15
Subjective	.53*	.29*	.33*
Objective	.53*	.16+	.22*
Assessment	.52*	.07	.18*
Plan	.43*	.26*	.17*
Clinician ID	.44*	.26*	.22*
Overall score	.59*	.26*	.29*

* $p < 0.01$
+ $p < 0.05$

Table 5
Spearman Correlation Coefficients:
Prevention Scores versus Record Documentation
For All 3 States and 3 Specialties Combined

Prevention

Documentation Category	Indicator			
	Blood Pressure	Flu Shot	Mammogram	Breast Exam
Date	.06	-.03	-.01	-.003
Subjective	.15*	.06	.15+	.29*
Objective	.10	.03	.19*	.29*
Assessment	.15+	.09	.11	.23*
Plan	.04	.05	.19*	.25*
Clinician ID	.04	.02	.18*	.24*
Overall score	.10	.09	.21*	.30*

* $p < 0.01$
+ $p < 0.05$

Table 6
Spearman Correlation Coefficients: Follow-up Scores versus Record Documentation For All 3 States and 3 Specialties Combined

Follow-up

Documentation Category	Indicator	
	Consult or Hospitalization	Significant findings
Date	-.02	.12
Subjective	.01	-.02
Objective	.01	.09
Assessment	-.06	.07
Plan	.04	.07
Clinician ID	.06	-.06
Overall score	.03	.01

* $p < 0.01$
+ $p < 0.05$

These data show that the direction of the association between documentation and performance score is as expected: positive. Better documentation does tend to be associated with a better performance score. However, the level of correlation is quite low in most cases. Thus there is no definitive link between documentation and performance scores.

Sample Size

A crucial question for all research endeavors is: how many cases should be reviewed in order to demonstrate the effect of an intervention? The following tables present the number of physicians and records per physician that are required to demonstrate certain levels of improvement in performance. We chose to present hypothetical sample sizes to demonstrate a very small improvement in performance, a 5 point improvement. The sample sizes for a 10 or 15 point improvement would be much smaller. We also chose a high level of power, or certainty that we could demonstrate the desired hypothetical difference. A lower level of power would mean smaller sample sizes. For example, if the power were dropped to 85%, then the sample size needed to demonstrate a 5 point improvement in diabetes monitoring would fall from 37 to 25.

Table 7
Sample Size Required to Demonstrate a
5 Point Improvement in
Old Diagnosis Monitoring Scores
95% power

Diagnosis	Number of Physicians Required, Assuming 20 patients/physician	Number of Patients per Physician Required, Assuming		
		20 Physicians	50 Physicians	100 Physicians
Diabetes	37	37	15	8
COPD	14	14	6	3
Congestive Heart Failure	10	10	4	2
Osteoarthritis	26	26	11	6
Hypertension	17	17	7	4
Ischemic Heart Disease	11	11	5	3

Table 8
Sample Size Required to Demonstrate a
5 Point Improvement in
Pre-existing Drug Monitoring Scores
95% Power

Diagnosis	Number of Physicians Required, Assuming 20 patients/physician	Number of Patients per Physician Required, Assuming		
		20 Physicians	50 Physicians	100 Physicians
ACE Inhibitors	54	54	22	11
Beta Blockers	25	25	10	5
Calcium Channel Blockers	46	46	19	10
Loop Diuretics	83	83	31	16
Insulin	85	85	34	17
Oral Hypoglycemics	157	157	63	32
Xanthine Derivative Drugs	50	50	20	10

Table 9
Sample Size Required to Demonstrate a
5 Point Improvement in
New Drug Monitoring Scores
95% Power

Diagnosis	Number of Physicians Required, Assuming 20 patients/physician	Number of Patients per Physician Required, Assuming		
		20 Physicians	50 Physicians	100 Physicians
ACE Inhibitors	55	55	22	11
Beta Blockers	48	48	19	10
Calcium Channel Blockers	68	68	28	14
Loop Diuretics	103	103	41	21
Insulin	138	138	56	28
Oral Hypoglycemics	128	128	51	26
Xanthine Derivative Drugs	54	54	22	11

Table 10
Sample Size Required to Demonstrate a
5 Percent Improvement in
HgbA1C Monitoring
95% Power

	Number of Physicians Required, Assuming 20 patients/physician	Number of Patients per Physician Required, Assuming		
		20 Physicians	50 Physicians	100 Physicians
Diabetes	48	48	20	10
Insulin	45	45	18	9
Oral Hypoglycemics	47	47	19	10

Table 11
Sample Size Required to Demonstrate a
5 Point Improvement in
Preventive Services
95% Power

	Number of Physicians Required, Assuming 20 patients/physician	Number of Patients per Physician Required, Assuming		
		20 Physicians	50 Physicians	100 Physicians
Influenza Immunization	96	96	39	20
Mammogram	90	90	36	18
Breast Exam	81	81	33	17

Table 12
Sample Size Required to Demonstrate a
5 Point Improvement in
New Diagnosis Evidence
95% Power

Diagnosis	Number of Physicians Required, Assuming 20 patients/physician	Number of Patients per Physician Required, Assuming		
		20 Physicians	50 Physicians	100 Physicians
COPD	68	68	28	14
Diabetes	92	92	37	17
Ischemic Heart Disease	126	126	51	26

Table 13
Sample Size Required to Demonstrate a
5 Point Improvement in
Appropriate Prescription
95% Power

Drug	Number of Physicians Required, Assuming 20 patients/physician	Number of Patients per Physician Required, Assuming		
		20 Physicians	50 Physicians	100 Physicians
Cephalosporin	95	95	38	19
Aminopenicillins	117	117	47	24
Erythromycin	65	65	26	13

Table 14
Sample Size Required to Demonstrate a
5 Point Improvement in
Response to an Abnormal Test
95% Power

Test	Number of Physicians Required, Assuming 20 patients/physician	Number of Patients per Physician Required, Assuming		
		20 Physicians	50 Physicians	100 Physicians
Blood Pressure	126	126	51	26
Hemoglobin	127	127	51	26
Potassium	87	87	35	18
Blood glucose	131	131	53	27
Creatinine	110	110	44	22
U/A Dipstick	130	130	52	26
Cholesterol	98	98	40	20

Summary and Recommendations

This chapter demonstrates the presence of variation in ambulatory care performance. The issue is, of course, what to do with the information. We present suggestions on how the DRRS data may be used to stimulate dialogue with physicians about practice variations. Some of the key questions to be posed to practicing physicians include:

- Do you find data believable and credible?
- Is this the appropriate standard for monitoring (test follow-up, etc.)? If not, how should the standard be revised?
- If the data are believable and the standard appropriate, what could have led to the finding?
- What are the barriers to improving practice?
- How can documentation of key information be facilitated?
- How can the office routine be organized to facilitate performance of key monitoring actions?

This chapter also shows that documentation and performance scores are only weakly correlated. Thus better documentation does not necessarily lead to better performance scores. This finding suggests that the variation demonstrated in this chapter is not related to quality of documentation, but rather to real variations in physician performance.

Finally, sample size calculations based on the variability in the present data show that the number of physicians required for an intervention study are reasonable.

Based on the data presented here, we recommend:

- Data showing practice variations should be used to initiate dialogue with physicians about the findings, their interpretation, and barriers to improvement.

Chapter 21

Costs of Implementing the DEMPAQ Record Review System (DRRS)

Abstract

Measuring the costs of introducing a new quality review system is crucial to inform decision makers who will need to weigh those costs and the benefits of broader implementation of the system. However, for a new program, it often is difficult to disentangle the costs of various activities including developing the review system, producing training materials, testing the validity and reliability of the system, reporting to the funding agency, and other ongoing activities of the organization.

In this chapter, our goal is to isolate the costs associated with the implementation of the DEMPAQ Record Review System (DRRS). By examining the costs of implementing DRRS, a computerized ambulatory medical record review system in three PROs, we hope to indicate the costs of replicating the system in other PROs.

This cost evaluation includes only the three areas of costs directly associated with the review system, the so-called marginal costs of DRRS. These are the costs that an organization could expect to face if it implemented a review system using their available review staff who had already been hired and trained as nurse reviewers for a PRO.

- Direct personnel costs for time reviewers spend reviewing records while signed onto the DRRS system.
- Personnel costs for time reviewers spend on miscellaneous activities related to the review.
- Administrative costs for tasks directly related to review such as obtaining medical records from physicians, making copies, filing and arranging the hard copies of the medical records, supervising reviewers, and training reviewers.

We do not consider here other important costs which the PRO would have borne even if DRRS were not implemented:

- PRO's fixed costs (e.g., building rental or general office administration);
- Recruitment or general training of new staff;
- Opportunity costs (benefits the program resources could achieve in alternate uses);
- Analysis of the results in order to provide feedback to physicians; and
- Dissemination of the results of the reviews to physicians through seminars or workshops.

We use as an example the DEMPAQ Record Review System (DRRS), a computerized review system based on data from samples of physician office records. This system was developed by the DEMPAQ project and was first implemented and tested in Maryland and then implemented in two other states, Iowa and Alabama. DRRS provides a means to measures ambulatory care performance using quality indicators based upon explicit criteria. Performance as measured by these indicators is tracked and compared over time and across physicians. The system runs on personal computers.

For each state, Table 1 on the following page shows the number of reviewers, months of review, number of cases included in the analysis, and the average time reviewers spent signed onto DRRS.

Table 1. Background Information

	Maryland	Iowa	Alabama
Months of review	2/5/92–4/17/92[1]	2/27/92–7/17/92	3/9/92–9/23/92
Number of reviewers	8[2]	3	2
Number of cases[3]	1,119	1,009	1,689
Average minutes per review (SD)	51.03 (46.18)	63.49 (67.20)	55.12 (47.39)

Methods

The PRO in each state provided information on the types of employees (supervisors, nurse reviewers, secretaries), average hourly salary and fringe benefits per type of employee, and time spent by each employee doing various activities related to the review (review time, copying records, training, etc.). As a measure of time signed onto the DRRS, we used minutes measured by the internal clock on the PC to generate a data set of reviewer ID, case ID, date reviewed, and minutes spent on the computer.

In all three states, the total number of hours that nurse reviewers reported as spending on review exceeded the sum of the minutes spent signed onto the computer. Therefore, we consider the cost of these as additional hours of review activity.

Results

We first report results based on the time reviewers were signed onto the computer system, DRRS, and then report results for the total marginal costs of implementing DRRS.

Time Spent Signed Onto DRRS

Broad Range of Review Times. On average, reviewers spend about an hour per case signed onto DRRS. Because medical records can vary widely in the amount of information about the patient and their treatment and the quality of how the information is recorded, we would expect to see a broad range of review times. Indeed, measured by the DRRS internal clock, the review times ranged from one minute to a maximum of 677 minutes for an extreme outlier in Iowa. The latter is probably a case of the reviewer neglecting to log off.

---

[1] While PQRS was initiated in Maryland on November 4, 1991, the version with an internal clock to track review times was only implemented after February 5, 1992.

[2] One reviewer (#1019), a supervisor, did only 14 reviews.

[3] Our initial examination of the data raised concerns that there were too many very short review times—under ten minutes. These could represent re-entry to edit the file or to complete a review. Such re-entries "wrote over" the original review time on the computer system. Therefore, we excluded cases with review times under ten minutes if there are:

- ❖ two different dates of review;
- ❖ two reviewer IDs;
- ❖ the number of visits exceeds the number of minutes for the review.

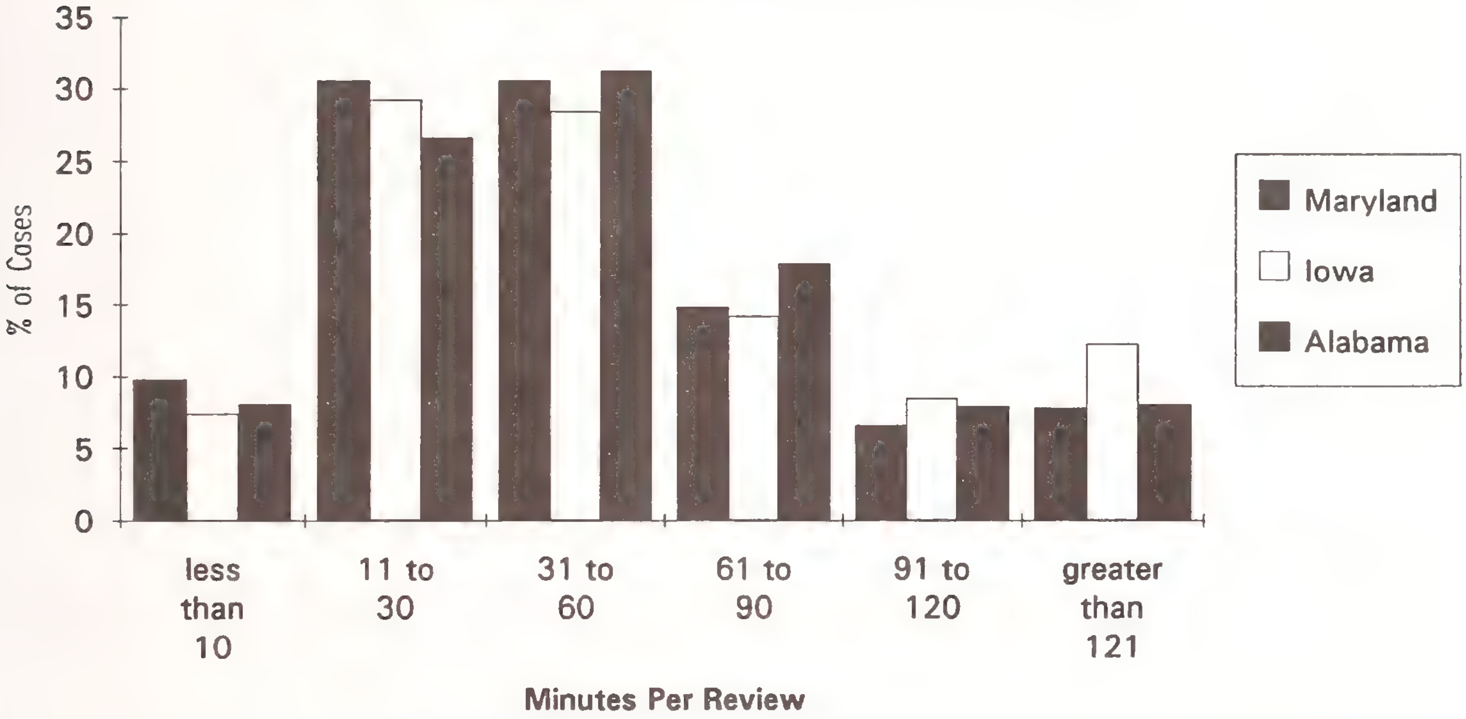
We kept cases which appear to have legitimately short review times:

- ❖ cases with "low" average RDA (LE 12);
- ❖ cases where more than three of a doctor's patients have short review times, probably indicating a problem with the doctor's charts.

This approach resulted in excluding 24 cases in Maryland, 97 cases in Alabama, and 75 cases in Iowa.

Figure 1 shows the distribution of review times for each state. Review times 10 minutes or under accounted for 8.1% of reviews in Alabama, 7.4% of reviews in Iowa, and 9.8% in Maryland. At the other end of the spectrum, review times two hours or over accounted for 8.1% of cases in Alabama, 12.3% of cases in Iowa, and 7.8% of cases in Maryland.

Figure 1. Frequencies of Minutes Per Review by State



Average DRRS Time per Reviewer. On the following page, table 2 shows the number of cases, average time spent signed onto DRRS, and average cost for time spent signed onto DRRS for each reviewer in each state. While some reviewers in Maryland showed statistically significantly faster average review times, we must be careful not to conclude that these reviewers were more efficient. A more likely explanation is that the more complex cases were not randomly distributed among reviewers although there was no systematic method used to assign cases to reviewers.

Table 2. Number of Cases, Minutes Signed onto DRRS, and Cost for Time Signed Onto DRRS

	Number of Cases	Minutes (SD.)	Cost (SD.)[5]
Maryland			
0985	299	46.24 (39.09)	$ 9.83 ($ 8.31)
0849	211	54.36 (40.52)	$11.55 ($ 8.61)
1025	194	40.57 (29.14)	$ 8.62 ($ 6.19)
1026	151	71.74 (61.42)	$15.24 ($13.05)
783	121	39.88 (35.51)	$ 8.47 ($ 7.55)
1027	72	90.94 (75.74)	$19.32 ($16.09)
0638	52	23.42 (15.22)	$ 4.98 ($ 3.23)
1019	14	32.29 (29.96)	$ 6.86 ($ 6.37)
Iowa**			
0940	427	57.22 (58.51)	$16.54 ($16.92)
0381	375	70.51 (73.87)*	$20.39 ($21.36)
427	203	63.87 (71.09)	$18.47 ($20.55)
Alabama***			
2000	851	55.48 (44.59)	$13.23 ($10.63)
1000	838	54.75 (50.09)	$13.06 ($11.95)

* This reviewer had the big 677 minute outlier. This average of 70.51 minutes is statistically significantly higher than review 0940 but not significantly different from reviewer 427.

*** The average times are not significantly different. For Iowa, the reviewer salary included fringe benefits ($17.35) while for Maryland ($12.75) and Alabama ($14.32), fringe benefits were calculated separately.

---

[5] Average salaries

Maryland		Iowa		Alabama	
Reviewer	$12.75	Reviewer	$17.35	Reviewer	$14.32
Supervisor	$14.67	Field staff	$18.61	Project Manager	$20.76
Clerical	$ 6.83	Programmer	$20.18	PC technician	$13.26
		Clerical	$12.96	Clerical	$11.73

Faster Review Times With Experience. Figure 2 below shows the average DRRS review time per reviewer per month. In Alabama, the average time declined from 68.52 minutes and 72.71 minutes per review during the first month of the review to 46.14 and 57.52 minutes by the last month of the review. In Iowa, review times also dropped. In Maryland no clear patterns emerged but the average review times are less stable because of small numbers per reviewer per month. This may also be attributable to the many changes made in the review system during Maryland's implementation, which did not occur in the other two states.

Figure 2. Average Review Time by Reviewer and Month, Alabama

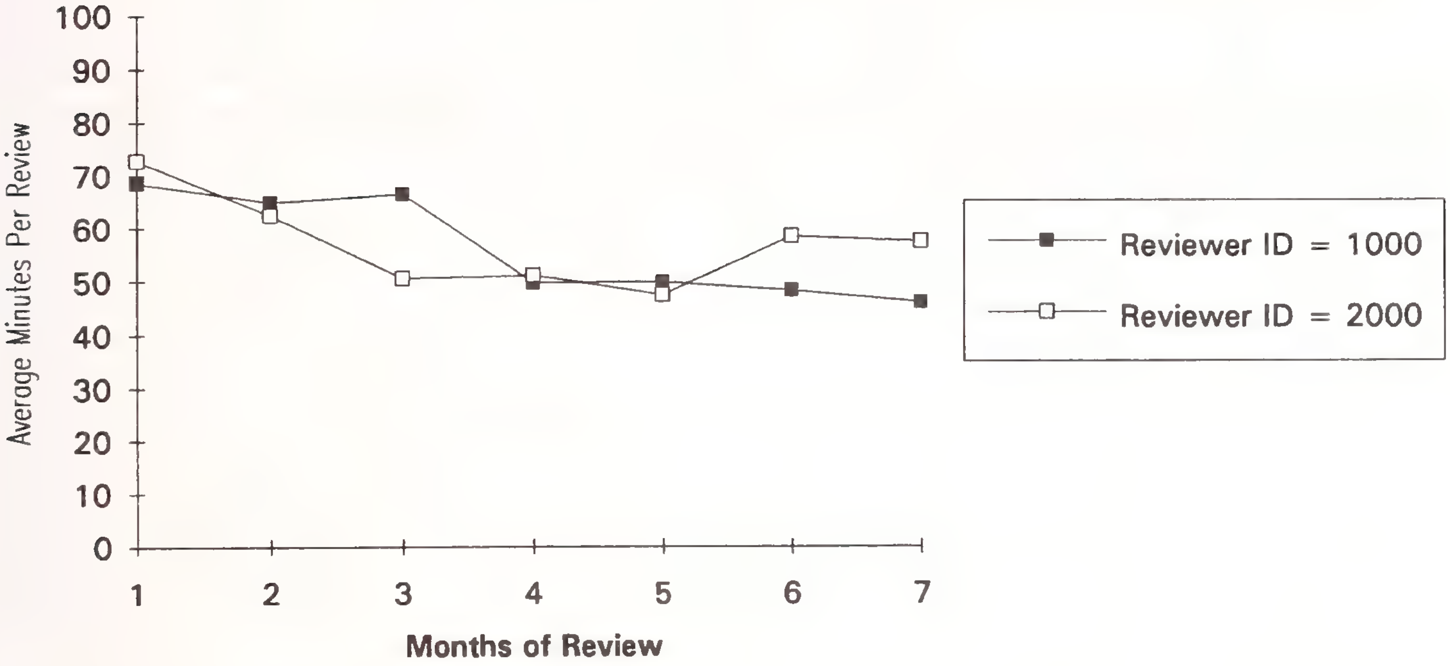
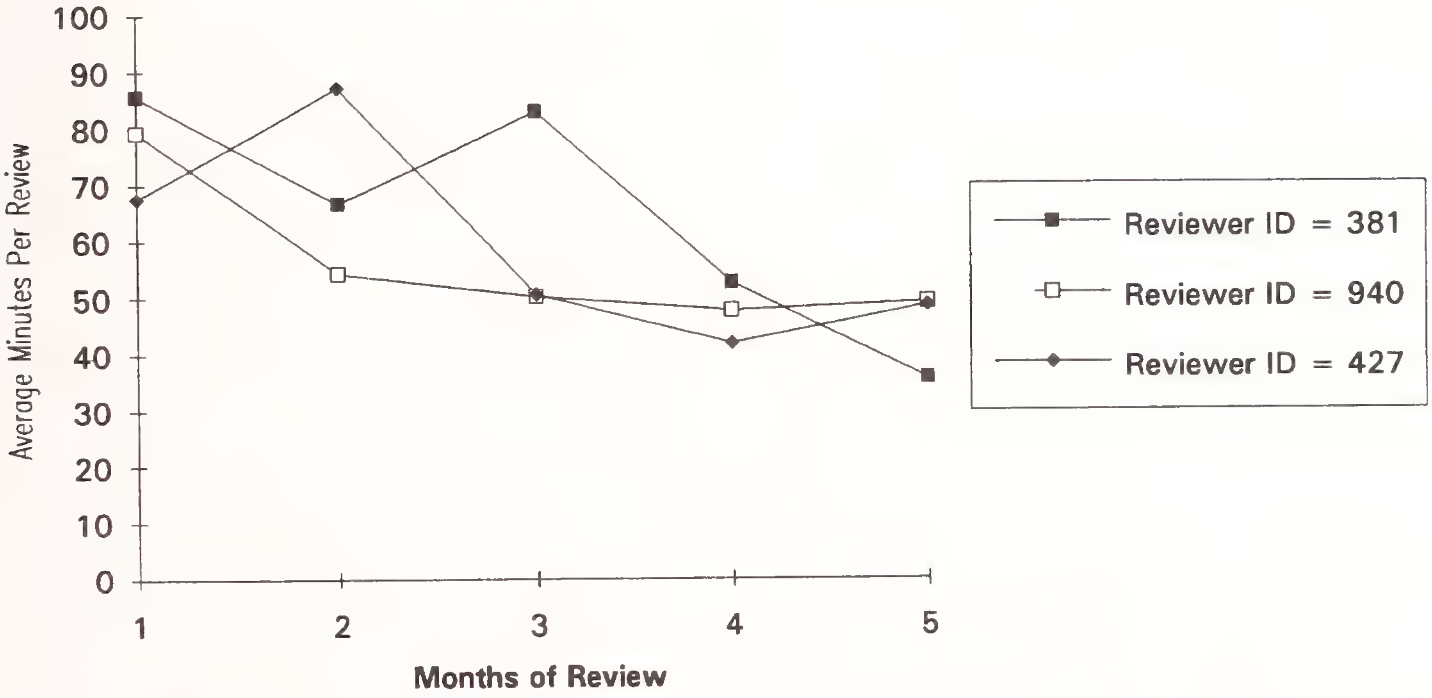


Figure 2. Average Review Time by Reviewer and Month, Iowa



Cost of Review

In addition to time signed onto the computer, the cost of review includes two additional components: reviewer's salary for additional time recorded as review activity; and the costs of other activities. These other activities include training in the DRRS system, copying records at physicians' offices, telephoning physicians, setting up offices and logs of records, and traveling to copy records in physicians' offices.

Table 3, on the next page, shows these additional costs for each state. Because each state used slightly different accounting methods to collect these costs, we should focus on several major findings:

- The average cost per review ranged from $35.05 in Alabama to $46.64 in Iowa to $73.34 in Maryland.
- The direct time signed onto the computer accounted for 14.7% of the review costs in Maryland, 39.8% in Iowa, and 35.6% in Alabama.

Note the original pilot testing and redesign of the DRRS was performed in Maryland. While we attempted to measure only the costs of the ongoing system, part of Maryland's cost per review should nonetheless be considered to represent the costs of development of the system.

Table 3. Overall Costs of DRRS

Time on DRRS	Maryland Hrs. ($)	Iowa Hrs. ($)	Alabama Hrs. ($)
Additional hours for review activity	1,159 ($14,781)	1,075 ($14,287)	1,557 ($22,296)
Medical Review Support (office setup, logs, etc.)	931 ($11,874)	397 ($5,273)	459 ($6,673)
Training	1,732 ($22,666)	714 ($8,964)	140 ($2,376)
Record copying at physician participants' offices	458 ($5,930)	132 ($2,168)	227 ($3,371)
Phone calls to physicians for participation	60 ($755)	131 ($1,734)	154 ($1,883)
Reimbursement to physicians for copying and postage	60 ($755)	not recorded	72 ($1,031)
Staff travel to copy office records	— ($5,377)	— ($4,845)	— ($9,696)
Fringe benefits	44 ($767)	— ($654)	— ($2,052)
Total	$19,160	$9,138	$9,914
Number of cases	4,444 ($82,064)	2,448 ($47,062)	2,609 ($59,292)
Total per case	1,119	1,009	1,689
	3.97 ($73.34)	2.43 ($46.64)	1.67 ($35.05)

Discussion
When a review system is designed, there is always a tradeoff between the length of time signed onto the computerized review system and the amount of information that can be collected. However, we conclude that the records should be used to code as much information as possible since the administrative costs of obtaining the medical records, copying them, etc., are substantial, in fact, over 60% of the cost of the review process.

With the exception of Maryland, where the costs may be higher because that state was the original state to develop and implement the DRRS, the average cost per review is under $50 per case. This suggests that implementing this review system in other state PROs might not be prohibitively costly. Moreover, in other types of organization, such as managed care plans, medical records might be readily available so that some of the administrative costs would be considerably less.

References

Chapter 1

1. U.S. Department of Health and Human Services, Public Health Service, Centers for Disease Control, National Center for Health Statistics. National ambulatory medical care survey 1989 summary. *Vital and Health Statistics,* series 13, No. 110. Hyattsville, Maryland, April 1992. DHHS Publication No. (PHS) 92-1771. See Table 3, p 23.

2. The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. *N Engl J Med.* 1993;329:977-86.

Chapter 2

1. Institute of Medicine. *Medicare: A Strategy for Quality Assurance* (Vol. 1). Kathleen N. Lohr, Ed., Washington, DC: National Academy of Sciences; 1990.

2. U.S. Bureau of the Census, *Statistical Abstract of the United States: 1992* (112th ed.). Washington, DC; 1992.

3. Palmer RH. The challenge and prospects for quality assessment and assurance in ambulatory care. *Inquiry.* 1988;25: 119-131.

4. USDHHS. Current estimates from the national health interview survey, 1991. *Vital and Health Statistics,* Series 10, Number 184. Washington, DC: DHHS; December 1992.

5. Sonnefeld S, Waldo P, Lemreux J, McKusick D. Projections of national health expenditures through the year 2000. *Health Care Financing Review.* 1991;13 (1): 1-27.

6. Palmer RH, Louis TA, HSCE L-N, et al. A randomized controlled trial of quality assurance in sixteen ambulatory care practices. *Med Care.* 1985;23:751-770.

7. McDonald CJ. Protocol-based computer reminders, the quality of care and the non-perfectibility of man. *N. Engl. J. Med.* 1976;295:1351-1355.

8. Barnett GO, et al. Quality assurance through automated monitoring and concurrent feedback using a computer-based medical information system. *Med Care.* 1978;16:962-970.

9. Berwick DM. Continuous improvement as an ideal in health care. *NEJM.* 1989;320:53-56.

Chapter 3

1. Palmer RH. The challenges and prospects for quality assessment and assurance in ambulatory care. *Inquiry.* 1988;25:119-131.

Chapter 4

1. American Diabetes Association, Committee on Professional Practice. Standards of medical care for patients with diabetes mellitus. *Diabetes Care.* 1989;12:365–8.

2. Daley J, Gertman PM, Delbance TL. Looking for quality in primary care physicians. *Health Affairs.* 1988;7:107–113.

3. Emmons DW, Wozniak GD. Physician profiling: studies on the socioeconomic environment of medicine. *American Medical Association.* 1993;9–14.

4. Findlay S. Profiling systems aim to eliminate second guessing of doctors. *Business and Health.* 1993;4:58–64.

5. Franks P, Nutting PA, Clancy CM. Health care reform, primary care, and the need for research. *JAMA.* 1993;270(12):1449–1453.

6. Garnick DW, Hendricks AM, Comstock CB, Pryor DB. *A Guide to Using Private Health Insurance Data for Medical Effectiveness Research.* Submitted for publication.

7. Hoy EW, Curtis RE, Rice T. Change and growth in managed care. *Health Affairs.* 1991;10(4):18–36.

8. Lasker RD, Shapiro DW, Tucker AM. Realizing the potential of practice pattern profiling. *Inquiry.* 1992;29(3):287–297.

9. Lawthers AG, Palmer RH, Edwards JE, Fowles J, Garnick DW, Weiner JP. *Developing and Evaluating Performance Measures for Ambulatory Care Quality, a Preliminary Report of the DEMPAQ Project.* The Joint Commission on Accreditation of Health Care Organizations. 1993.

10. Leatherman S, Peterson E, Heinen L, Quam L. Quality screening and management using claims data in a managed care setting. *Quality Review Bulletin.* 1991; :349–359.

11. McIlrath S. Medicare asks: what's your specialty? *AMA News.* 2/10/1992;3:25.

12. McNeil BJ, Pederson SH, Gastonis C. Current issues in profiling quality of care. *Inquiry* 1992;29(3):298–307.

13. Nathanson P, Noether M, Ozminkowski RJ, Smith KM, Raney BE, Mickey D, Hawley PM. Using claims data to select primary care physicians for a managed care network. 1993. In progress.

14. Palmer RH. *Inquiry.* 1988.

15. Parente ST, Weiner JP, Richards T, Chandler P, Garnick DW, Fowles J. Methods for the development of a quality improvement database using insurance data: an application of Medicare's national claims history file. 1993. In progress.

16. Physician Payment Review Commission. Conference on Profiling. No. 92–2. Washington D.C.: Physician Payment Review Commission. 1992.

17. Ricci JA. Monitoring the new evaluation and management codes. *Medicare Special Bulletin.* 1992.

18. Shapiro DW, Lasker RD, Bindman AB, Lee PR. Containing costs while improving quality of care: the role of profiling and practice guidelines. *Annual Review of Public Health.* 1993;4:219–241.

19. Robinson ML. Medicare carriers to begin profiling physician practice patterns. *Report on Medical Guidelines & Outcomes Research.* 1990;1(4):4–5,12.

20. Stephenson GM. Blues plan uses profiling data to assess providers. *Report on Medical Guidelines & Outcomes Research.* 1992;13(8):6–7.

21. Weiner JP, Powe NR, Steinwacks DM, Dent. Applying insurance claims data to assess quality of care: a compilation of potential indicators. *Quality Review Bulletin.* 1990;16(12):424–438.

22. Weiner JP. Ambulatory case-mix methodologies: application to primary care research. in Hibbard H, Nutting PA, Grady M, eds. *Primary Care Research: Theory and Methods.* U.S. Department of Health and Human Services.

23. Abramson JH. *Survey Methods in Community Medicine.* New York: Churchill Livingstone. 1990.

Chapter 5

1. Anderson G, Steinberg EP, Whittle J, Powe NR, Antebi S, and Herbert R. Development of clinical and economic prognoses from medical claims data. *JAMA.* 1990;253:967–972.

2. Berwick DM. Continuous improvement as an ideal in health care. *New England Journal of Medicine.* 1989;320(1):53–56.

3. Washington, D.C.: U.S. Government Printing Office. *Federal Register.* June 5, 1991.

4. Fisher ES, Malenka DJ, Wennberg JE, and Roos NP. Technology assessment using insurance claims: example of prostatectomy. *International Journal of Technology Assessment in Health Care.* Vol. 6.1990;194–202.

5. Garnick DW, Hendricks AM, Comstock CB , and Pryor DB. A guide to using private insurance data for medical effectiveness research. In press. 1993.

6. Health Care Financing Administration. *Program Statistics. Medicare and Medicaid Data Book, 1988.* Baltimore: Health Care Financing Administration. 1988.

7. Hornbrook MC, Hurtado, AV and John, RE. Health care episodes: definition, measurement and use *Medical Care Review.* Vol. 42, No. 2. 1985:163–218.

8. Iezzoni, L. "Black box" medical information systems: a technology needing assessment. *JAMA.* 1991;265(22): 3006–3007.

9. Jencks SF and Wilensky GR. The health care quality improvement initiative. *JAMA.* 1992;268(7):900–903.

10. Lave, JR. et al. Costing medical care: using Medicare administrative data. May 1993. In press.

11. Lawthers AG, Palmer RH, Edwards JE, Fowles J, Garnick DW, and Weiner, JP. Developing and evaluating performance measures for ambulatory care quality. *Quality Review Bulletin.* 1993. In press.

12. Lohr, KN. Use of insurance claims data in measuring quality of care. *International Journal of Technology Assessment in Health Care.* 1990;6(2):263–71.

13. McDonald, CJ and Hui SL. The analysis of humongous databases: problems and pitfalls. *Statistics in Medicine.* 1991;10(4):511–518.

14. McCall, N. Personal Communication. Health Care Financing Administration. 1992.

15. Newhouse, JP. Medical effectiveness research data methods: summary and reactions. In Grady ML, Schwartz, HA, eds. *Medical Effectiveness Research Data Methods.* Rockville, M.D.: Agency for Health Care Policy Research. 1992.

16. Parente ST, Weiner JP, Fowles J, Lawthers AG, Palmer RH, and Garnick DW. Incorporating primary care source (PCS) assignment into medical practice variations analysis. *Medical Decision Making*. October–December 1992.

17. Warren JL, Babish JD, and Nicholson G. Use and linking of Medicare databases in creating episode of care file. Presented at the American Public Health Association 1990 Annual Meeting.

18. Weiner JP, Powe NR, Steinwachs DM, Dent G. Applying insurance claims data to assess quality of care: a compilation of potential indicators. *Quality Review Bulletin*. 1990;16(12):424–438.

19. Weiner JP, Starfield BH, Steinwachs DM and Mumford L. Development of a population-oriented measure of ambulatory care case-mix. *Medical Care*. 1991;29(5):452–472.

Chapter 9

1. Roos LL, Sharp SM, Wajda A. Assessing data quality: a computerized approach. *Soc. Sci. Med.* 1989;28:175.

2. Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP. Accuracy of diagnostic coding for medicare patients under the prospective payment dystem. *NEJM.* 1988;318:352.

3. Lloyd SS, Rissing JP. Physician and coding errors in patient records. *JAMA*.1985;254:1330.

4. Feigl P, Glaefke G, Ford L, Diehr P, Chu J. Studying patterns of cancer care: how useful is the medical record? *AJPH*.1988;78:526.

5. Iezzoni LI, Burnside S, Sickles L, Moskowitz MA, Sawitz E, Levine PA. Coding of acute myocardial infarction: clinical and policy implications. *Ann Int Med.* 1988;109:745.

6. Demlo LK, Campbell PM, Brown SS. Reliability of information abstracted from patients' medical records. *Med Care*. 1978;16:995.

7. Wennberg JE, Roos N, Sola L, Schori A, Jaffe R. Use of claims data systems to evaluate health care outcomes: mortality and reoperation following prostatectomy. *JAMA*. 1987;257:933.

8. Roos LL, Cageorge SM, Austen E, Lohr KN. Using computers to identify complications after surgery. *AJPH*. 1985;75:1288.

9. Roos LL, Fischer ES, Sharp SM, Newhouse JP, Anderson G, Bubolz TA. Postsurgical mortality in Manitoba and New England. *JAMA*. 1990;263:2453.

10. Luft HS, Hunt SS. Evaluating individual hospital quality through outcome statistics. *JAMA*. 1986;255:2780.

11. Anderson G, Steinberg EP, Whittle J, Powe NR, Antebi S, Herbert R. Development of clinical and economic prognoses from Medicare claims data. *JAMA*. 1990;263:967.

12. Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T. Comorbidities, complications and coding bias: does the number of diagnosis codes matter in predicting in-hospital mortality? *JAMA*. 1992;267:2197.

13. Maronde RF, Chan LS, Larsen FJ, Strandberg LR, Laventurier MF, Sullivan SR. Underutilization of antihypertensive drugs and associated hospitalization. *Med Care*. 1989;27:1159.

14. Palmer RH. The challenges and prospects for quality assessment and assurance in ambulatory care. *Inquiry*. 1988;25:119.

15. Horner RD, Paris JA, Purvis JR, Lawler FH. Accuracy of patient encounter and billing information in ambulatory care. *J Fam Prac.* 1991;33:593.

16. Ellis LBM, Krogh C, Werth G. Noise and validity in a practice-derived database. In Miller RA, ed. *Proceedings of the Fourteenth Annual Symposium on Computer Applications in Medical Care.* Washington, DC: Institute of Electrical and Electronics Engineers. 1990:271.

17. Lurie N, Popkin M, Dysken M, Moscovice I, Finch M. Accuracy of diagnoses of schizophrenia in Medicaid claims. *Hosp Comm Psych.* 1992;43:69.

18. Roos LL, Roos NP, Cageorge SM, Nicol JP. How good are the data? reliability of one health care data bank. *Med Care.* 1982;20:266.

19. Leatherman S, Peterson E, Heinen L, Quam L. Quality screening and management using claims data in a managed care setting. *QRB.* 1991;349.

20. Buchsbaum DG, Boling P, Groh M. Residents' underdocumentation in elderly patients' records of prescriptions for benzodiazepine. *J Med Educ.* 1987;62:438.

21. Moran MT, Wiser TH, Nanda J, Gross H. Measuring medical residents' chart-documentation practices. *J Med Educ.* 1988;63:859.

22. Romm FJ, Putnam SM. The validity of the medical record. *Med Care.* 1981;19:310.

23. Quam L, Ellis LBM, Venus P, Clouse J, Taylor CG, Leatherman S. Using claims data for epidemiologic research: the concordance of claims-based criteria with the medical record and patient survey for identifying a hypertensive population. *Med Care.* 1993;31:498.

24. Coy JA, Fanta CM, Finkel AJ et al. eds. *Physicians' Current Procedural Terminology,* 4th ed. Chicago: American Medical Association. 1988.

25. Fleiss JL. *Statistical Methods for Rates and Proportions.* 2nd ed. New York: John Wiley & Sons. 1981.

26. Lawthers AG, Palmer RH, Edwards JE, Fowles J, Garnick DW, Weiner JP. Developing and evaluating performance measures for ambulatory care quality: a preliminary report of the DEMPAQ project. *Joint Commission Journal for Quality Improvement* (in press).

27. PMIC. *International Classification of Diseases;* 9th revision, 3rd edition: clinical modification, vols 1, 2 &3. Los Angeles: Practice Management Information Corporation. 1989.

28. Rosner B. *Fundamentals of Biostatistics.* 2nd ed. Boston Mass: Duxbury Press. 1986.

Chapter 10

1. Abramson, JH. *Survey Methods in Community Medicine.* New York: Churchill Livingstone. 1990.

2. Abrams WB, Berkow R, eds. *The Merck Manual of Geriatrics.* Rahway, NJ: Merck & Co. 1990.

3. Branch WT. *Office Practice of Medicine,* 2nd ed. Philadelphia: WB Saunders. 1987.

4. Brechner RJ et al. Ophthalmic examination among adults with diagnosed diabetes mellitus. *JAMA.* 1993;270(14):1708-1713.

5. Cassel CK, Riesenberg DE, Sorensen LB, et al. *Geriatric Medicine,* 2nd ed. New York: Springer-Verlag. 1990.

6. Cohn JN. Current therapy of the failing heart. *Circulation.* 1988;78:1099-1107.

7. Cohn JN. Prognostic factors in heart failure: poverty amidst a wealth of variables. *J Am Coll Cardiol.* 1989;14:571-2.

8. Fanta CH, Ingram RH. Chronic obstructive diseases of the lung. Ch. III. In: *Scientific American Medicine.* Section 14. Respiratory medicine. New York: Scientific American, 1988.

9. Fleg JL, Hinton PC, Lakatta EG, et al. Physician utilization of laboratory procedures to monitor patients with congestive heart failure. *Arch Intern Med.* 1989;149:393-6.

10. Goroll AH, May LA, Mulley A, eds. *Primary Care Medicine,* 2nd ed. Philadelphia: JB Lippincott, 1981.

11. Hayward RSA, Steinberg EP, Ford DE, et al. Preventive care guidelines: 1991. *Ann Intern Med.* 1991;114:758-83.

12. Johanson WG, Harris MS. *COPD Reference Guide,* 2nd ed. American Board of Family Practice. 1988.

13. Kannel WB, Belanger AJ. Epidemiology of heart failure. *Am Heart J.* 1991;121(no. 3, part 1):951-7.

14. Levine GI, Nong A. *Coronary Artery Disease Reference Guide*, 3rd ed. American Board of Family Practice. 1990.

15. National Center for Health Statistics. Detailed diagnoses and procedures, national hospital discharge survey, 1989. *Vital and Health Statistics,* Series 13, no. 108. DHHS publication no. (PHS) 91-1769. Hyattsville, MD: National Center for Health Statistics. 1991.

16. Nong A, Reams CJ. *Chronic Heart Failure Reference Guide,* 3rd ed. American Board of Family Practice. 1990.

17. US Preventive Services Task Force. *Guide to Clinical Preventive Services.* Baltimore: Williams and Wilkins. 1989.

18. Wilson JD, Braunwald E, Isselbachen KJ, et al., eds. *Harrison's Principles of Internal Medicine,* 12th ed. New York: McGraw-Hill. 1991.

Chapter 11

1. Romm FJ, Putnam SM. The validity of the medical record. *Medical Care.* 1981;19(3):310-315.

2. *Paradox, Relational Database.* Version 3.5. Scotts Valley, CA: Borland International; 1990.

3. Charles, Wilber & Kennedy. Physicians' self-reports of reactions to malpractice litigation. *Am J of Psychiatry.* 1984;41;563-565.

4. Charles, Wilbert & Franke. Sued and nonsued physicians' self-reported reactions to malpractice litigation. *Am J. Psychiatry;*1985:437-440.

5. Peters, Nord, and Woodson. An empirical analysis of the medical and legal professions' experiences and perceptions of medical and legal malpractice. *J Law Reform.* 1986:601-636.

6. *SAS Language Guide for Personal Computers.* Release 6.03. Cary, North Carolina: SAS Institute Inc. 1988

Chapter 13

1. Abramson JH. *Survey Methods in Community Medicine.* New York: Churchill Livingstone. 1990.

2. Fleiss, JL. *Statistical Methods for Rates and Proportions.* New York: John Wiley and Sons. 1981.

Chapter 14

1. Abramson, JH. *Survey Methods in Community Medicine.* New York: Churchill Livingstone. 1990.

Chapter 19

1. Weiner J et al. Developing and testing a population-oriented measure of ambulatory care case-mix. *Care.* May 1991.

Chapter 21

1. Thompson MS, Palmer RH, Rothrock JK, Strain R, Brachman IH, Wright EA. The cost of quality assurance in medicine. *Evaluation and the Health Professions.* 1983;6(3)283-307.

PROPERTY OF
AUG 21 2000
HCFA LIBRARY

CMS LIBRARY
3 8095 00014518 1